

DOCUMENT RESUME

ED 034 730

SP 003 405

AUTHOR Beatty, Walcott H., Ed.
 TITLE Improving Educational Assessment & An Inventory of Measures of Affective Behavior.
 INSTITUTION Association for Supervision and Curriculum Development, Washington, D.C.
 PUB DATE 69
 NOTE 174p.
 AVAILABLE FROM Publications-Sales Section, National Education Association, 1201 16th Street, N.W., Washington, D.C. 20036 (Stock No.: 611-17804, \$3.00)

EDRS PRICE MF-\$0.75 HC Not Available from EDRS.
 DESCRIPTORS *Affective Behavior, Attitude Tests, Classroom Observation Techniques, Creativity, Curriculum Evaluation, *Educational Testing, Educational Theories, *Evaluation, Interaction, Measurement Goals, *Measurement Instruments, Motivation, Personality Tests, Program Evaluation, Readiness, Self Concept

ABSTRACT

The first half of this publication consists of four papers presented at a 1967 working conference intended to foster the development of a theory of educational assessment. Topics discussed in "The Purposes of Assessment" by Ralph W. Tyler include assessment for diagnosis, for individual guidance, for college admissions and placement, and assessment of pupil readiness, of innovations, and of learning materials and procedures. In "Language, Rationality, and Assessment," Robert E. Stake's topics include curriculum evaluation, congruence and contingency, generalizability of findings, and rationalism and empiricism. "Evaluation as Enlightenment for Decision Making" by Daniel L. Stufflebeam includes sections on the state of the art in educational evaluation and on the nature of evaluation. The final essay by Walcott H. Beatty, "Emotion: The Missing Link in Education," focuses on self-concept, motivation, and learning and the promotion of affective development. The second half of the book is an annotated resource list of 133 instruments already developed or under development for measuring eight different categories of affective behavior. The categories and number of instruments reviewed in each are Attitude (10), Creativity (7), Interaction (15), Miscellaneous (19), Motivation (27), Personality (23), Readiness (3), and Self-Concept (29). The measures are indexed by author, title, and abbreviation. (JS)

ED034730

PROCESS WITH MICROFICHE AND
PUBLISHER'S PRICES. MICRO-
FICHE REPRODUCTION ONLY.

*Improving
Educational
Assessment*

&

*An Inventory
of Measures
of Affective
Behavior*

Prepared by the ASCD Commission on
Assessment of Educational Outcomes

WALCOTT H. BEATTY, Chairman and Editor

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE
PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION
POSITION OR POLICY.

Association for Supervision and Curriculum Development, NEA
1201 Sixteenth Street, N.W., Washington, D.C. 20036

SP003405

Permission to reproduce this copyrighted work has been granted to the Educational Resources Information Center (ERIC) and to the organization operating under contract with the Office of Education to reproduce documents included in the ERIC system by means of microfiche only, but this right is not conferred to any users of the microfiche received from the ERIC Document Reproduction Service. Further reproduction of any part requires permission of the copyright owner.

Copyright © 1969 by the
Association for Supervision and Curriculum Development, NEA

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the publisher.

Price: \$3.00

NEA Stock No.: 611-17804

The materials printed herein are the expressions of the writers. They do not represent endorsement by, nor a statement of policy of, the Association.

Library of Congress Catalog Card Number: 79-101378

Contents

iv	Acknowledgments
v	Foreword <i>Alexander Frazier</i>
vi	Preface <i>Walcott H. Beatty</i>
	Section I. Improving Educational Assessment
2	The Purposes of Assessment <i>Ralph W. Tyler</i>
14	Language, Rationality, and Assessment <i>Robert E. Stake</i>
41	Evaluation as Enlightenment for Decision Making <i>Daniel L. Stufflebeam</i>
74	Emotion: The Missing Link in Education <i>Walcott H. Beatty</i>
	Section II. An Inventory of Measures of Affective Behavior <i>Donald J. Dowd, Sarah C. West</i>
	CATEGORIES:
90	Attitude Scales
95	Creativity
99	Interaction
107	Miscellaneous
116	Motivation
129	Personality
141	Readiness
143	Self Concept
	INDEXES TO THE INVENTORY:
159	By Authors
160	By Titles of Measures
162	By Abbreviations Associated with Titles
164	Contributors
164	Members of the ASCD Council on Assessment of Educational Outcomes

Acknowledgments

FINAL editing of the manuscript and production of this book were the responsibility of Robert R. Leeper, Associate Secretary and Editor, ASCD Publications. Technical production was handled by Mary Albert O'Neill, Lana G. Pipes, Nancy Olson, and Karen T. Brakke. Nancy Olson and Lana G. Pipes compiled the indexes.

Foreword

In education, as in every other field of work, the quality of the decisions made is always limited by the data we have at hand. Now we are newly aware that the information we have had has never been good enough. For one thing it has never been sharply and dependably accurate, even as to our more ordinary subject-matter goals. More important, it has rarely accounted for the broader scope of fundamental objectives or been designed to serve the wide variety of purposes to which evaluation is germane.

In part our new awareness may stem from being called more sharply to account for the scope and outcomes of our decisions. In part, perhaps, it results from the broadened base of persons involved in educational planning and replanning. Anyway, we know we have to have richer information—as well as wider ambitions—if we are to improve education in ways that really do count.

Thus, this publication should be most useful. Its first section consists of papers which analyze the system we have for collecting and using data—and which propose extensions of the system to catch new purposes and new dimensions. Its second section is a comprehensive, well annotated resource list of devices already developed or under development which may enable us to tap into aspects of human behavior which we have always found hard to get at in anything like objective terms. Both sections should add something of value in preservice and in-service education as well as in study at the graduate level.

The whole Council responsible for this work deserves the Association's thanks. We are particularly indebted to its chairman, Walcott H. Beatty, for his careful work in assembling and editing the papers.

October 1969

ALEXANDER FRAZIER, *President*
Association for Supervision
and Curriculum Development, NEA

Preface

THE Council on Assessment of Educational Outcomes was established in May 1965 by the Executive Committee of the Association for Supervision and Curriculum Development. The original charge to the Council was:

1. To determine appropriate objectives of assessment for individuals, teachers, the school system, and governmental agencies;
2. To analyze critically existing assessment procedures in the light of such objectives;
3. To explore broader aspects of assessment relating to individual growth and educational objectives, including the place of local, economic, and social conditions and values and the effects of assessment processes on students, teachers, and programs; and
4. To stimulate new approaches to appraisal designed to tap the broadest possible spectrum of individual learning and growth.

The Council was created by ASCD in response to concern about the plans of the Exploratory Committee on Assessing the Progress of Education operating under a grant from the Carnegie Corporation to design a national assessment of the educational attainments of the American people. Walcott H. Beatty was asked to be chairman of the Council and additional appointments of Earl C. Kelley and Jack R. Frymier completed the membership. In 1966 Donald J. Dowd was added to the Council, and in 1967 Earl C. Kelley resigned from its membership.

At its first meeting in October 1965, the Council was asked to draft a set of guidelines for national assessment of educational outcomes. This draft was later revised and issued as a statement by the ASCD Executive Committee in the January 1966 issue of the *ASCD News Exchange*. During a series of meetings held in

1966 the Council decided that its original charge was too broad to provide clear guidance; rather, it held that three kinds of activities should be undertaken. Its first project would be an attempt to determine how effectively educational objectives can be assessed at the present time. A second activity would be an attempt to develop guidelines for evaluation of large projects such as those funded by the federal government. The discussions of the Council led to a conclusion that current measurement theory is inadequate for purposes of educational assessment. The Council proposed that a working conference on assessment theory be developed.

Work was begun immediately to carry forward these activities and to develop a more specific charge to the Council. The following revised charge was approved by the Executive Committee of ASCD in May 1967:

1. To foster the development of a theory of educational assessment;
2. To define problems in carrying out effective assessment and to develop means for coping with these problems;
3. To review existing instruments in the area of self concept; to provide information relating to measures of self concept and other noncognitive assessment instruments; and
4. To recommend policy positions on assessment to the ASCD Executive Committee.

The gathering of information about instruments to measure self concept and other noncognitive factors was begun and detailed planning for a working conference on assessment theory was completed. The outcomes of these two projects are presented in this publication.

The Working Conference on Assessment Theory

In May 1967 the conference plan was approved by the ASCD Executive Committee and was scheduled for January 18-20, 1968, at Sarasota, Florida. The conference was funded by ASCD with the help of a grant from the National Institute of Mental Health. The rationale for the conference was stated as follows:

Measurement theory which currently exists appears to be much too restrictive to serve the purposes of evaluating the effectiveness of educational procedures. Such assessment is not really concerned with the

mean level of achievement of children in various grades, subjects, or schools, but rather with the assessment of how effectively current curricula and teaching methods accomplish the objectives stated for a course, program, or total school.

Such an approach might, for example, require the operationalizing of objectives, the development of breakdowns of student populations according to significant variables, the determination of school activities which are relevant to school objectives, the testing of a variety of curricula and methodological approaches, etc. It is the purpose of the conference to clarify thinking in these areas.

Four speakers were invited to address the conference. Approximately 40 persons active in educational assessment at federal, state, and local levels and at universities were also invited to attend. After each speaker presented his ideas the participants discussed them in small groups. There was also a final panel of participants to share the questions and new thinking which had emerged from the conference.

At its subsequent meeting, the Council decided that the working conference had made some contribution toward the first two tasks assigned to the group and that the information gathered by the Council on noncognitive assessment instruments met the third charge. The Council decided to request the Publications Committee of ASCD to recommend publication of a booklet containing the invited addresses presented at the conference¹ and an annotated bibliography of noncognitive assessment instruments being developed nationwide for the measurement of learning and growth.

The Council also concluded that, while there is yet much to be done in the area of educational assessment, the members of this ASCD group had made their maximum contribution and that a new group with new ideas should be formed to carry the work forward. The Council therefore recommended that it be dissolved concurrent with the publication of its work.

WALCOTT H. BEATTY, *Chairman*
DONALD J. DOWD
JACK R. FRYMIER

¹ Due to the lack of availability of a treatment of the affective domain at the conference on assessment, the paper "Emotion: The Missing Link in Education," by Walcott H. Beatty is included here although it was not presented at the conference.

Section I

*Improving
Educational
Assessment*

The Purposes of Assessment

RALPH W. TYLER

IN THE past two decades, educational assessment, evaluation, and appraisal have undergone profound changes. One fundamental development is the range of new uses for measurement and evaluation, including such pupil services as guidance, admission, and placement; the awarding of scholarships; diagnosis of student learning and development; the appraisal of new programs, courses, instructional procedures, materials, equipment; the management and guidance of programs in the schools; and the assessment of progress in education for public understanding and public policy.

These changes have arisen both from the changing social situation and from the development of new knowledge and technology in agriculture, industry, defense, commerce, and the health services, shifting the requirements of many occupations from physical strength and manual dexterity to intellectual activity and social sensitivity and skills. At present, the great employment opportunities are in science, engineering, education, the health services, recreation services, social services, management, and accounting.

In our time, the role of the school has shifted from that of selecting a small percent of the pupils for more advanced education while the others dropped out and went to work to that of reaching every child effectively to enable him to go on learning far beyond the expected level of 25 years ago. The task of the college is no longer to find the favored few but to identify a wide range of potential talents and to help each student to achieve his potential, both for his own self-realization and to meet the ever

increasing demands of a complex technological society. This changing situation requires new instruments and new procedures of educational evaluation, assessment, and measurement.

New knowledge and technology in education are influencing evaluation. For example, the findings of many recent studies have clarified the need for evaluating the powerful effects of the student's home, culture, and community environment upon his learning. A series of investigations, like that of Newcomb and Coleman, has shown the strong influence of peer group attitudes, practices, and interests upon the learning of its members, thus indicating the need for evaluating the nature, direction, and amount of peer group influences in developing effective school programs. These significant studies have brought about new conceptions and new practices in several areas of educational assessment and evaluation.

Pupil Readiness

One purpose of assessment of the individual pupil is to determine his readiness to pursue the next step of learning. We are so tied to the kind of achievement test which is focused on the middle level of difficulty that we have not examined what kinds of assessment the teacher needs to determine whether the curriculum is really sequential and whether the student has mastered a particular set of basic concepts and is ready to move on to the next stage in the process. That kind of test—a so-called mastery test—is one which samples the concepts which are basic to the next step in the sequence. Such a test includes samples of exercises at the minimum level of difficulty; that is, the most basic understanding.

Consider the program of Individually Prescribed Instruction which the University of Pittsburgh's Research and Development Center developed for the Oakleaf School. For assessment purposes, this program accepts a mastery of 85 percent of these exercises, recognizing that there will be a certain number of errors which are due to chance factors. Evidence shows that one of the reasons why youngsters fail is that the teacher paces instruction in the wrong way, aiming at the understanding of the average in the class. The teacher then moves on; the youngster at the lower end, not having understood the first stage, cannot understand the next stage. Finally, the cumulative effect of not understanding and not having developed the basic skills makes the youngster at the lower end so far behind that he gives up trying. The kind of assessment

which can be useful to the teacher is a set of exercises which samples the basic ideas at the simplest level, which is quite a different thing from the common achievement test that we have been using.

Diagnosis

Another basic purpose for assessment by the teacher is diagnosis or, as Marion Jenkins referred to it in her paper before the Third International Curriculum Conference, "troubleshooting." Jenkins points out that the concept of diagnosis comes from the field of medicine where known diseases can be identified. In education we do not always know whether there is a disease or a set of criteria that can be identified. We only know that the pupil is in difficulty—in that sense "troubleshooting" may be a better word. Whether we call it "diagnosis" or "troubleshooting," such a procedure requires special criteria for assessment. It requires an assessment of the student's environment in order to evaluate his potential success in moving ahead—home environment, language used in the home, types of behavior valued by the student's peer group, and interests and previous experience. However, we have not usually thought in this way when we have talked about giving diagnostic tests.

The problem of diagnosis and troubleshooting in group instruction requires a different kind of assessment. Thelen has pointed out in his very provocative monographs on group instruction that the group can become a learning organism, stimulating, providing meaning, and helping to reinforce learning. In groups, diagnosis demands a study of how the group is operating and how well it is succeeding. Some of the diagnostic procedures that we are accustomed to think about in group dynamics or in the activities of the National Training Laboratories or in the laboratories of persons such as Lippitt and Thelen are useful here. Diagnosis to determine the extent to which the group is operating effectively as a mechanism for encouraging, stimulating, and directing learning is something that has not usually been included in our approach to assessment.

Individual Guidance

Another purpose of assessment is to provide individual guidance. We have failed to analyze carefully enough what kind of

assessment information and procedures we need to guide the individual. We have based much of our effort on prediction; for example we have analyzed what proportion of persons with a particular pattern of test scores and other characteristics have gone into a certain field or have been successful in a field and then compared this pattern with our evaluation of a particular child.

A more dynamic notion of what happens in educational and vocational guidance is one of a continuing development. We need information about the student's own plans—how he has been planning, what factors he has considered in thinking about his future, and what stage he thinks is the next step. The problem of the guidance person is to identify enough of the student's background to help him take the next step for his own exploration, rather than to say, "Well, now we know he is going to be a good teacher because he has this particular pattern." This type of guidance requires information about the person's background and a dynamic picture of his present stage in the exploration of his own identity and his opportunities. Predictions are useful for a group but do not help the student very much with his problems.

Assessing Innovations

Innovations in curriculum development require appropriate assessment. At least two stages of assessment can be identified in the development of curriculum materials, procedures, tools, and media. One stage is the detailed examination of the total curriculum to see that each part is consistent with the general aim. For example, 20 years ago, the Commissioner of Education of the City of New York asked a committee of three to evaluate the activity school programs in the elementary schools. The members of the committee were George Stoddard, then Dean of the Graduate School of the State University of Iowa; Paul Rankin, then Director of Research at Detroit, and I. The original design of these 18 schools was to provide a kind of treatment; the other schools represented the control. The commissioner wanted to find out what kinds of educational results were being attained by the two.

We sat a good many days in the classrooms in these schools. We involved ourselves rather deeply in the philosophy of the activity schools, and we identified some 61 characteristics that were supposed to be a part of activity schools, including the involvement

of pupils in the planning of educational aims and so on. We found tremendous variation among the 18 schools. There were some classrooms in which almost every one of these 61 things were being done; they were really activity schools in the sense of the original plan. In other classrooms we found that anywhere from half to almost none of the characteristics were present. All 18 schools called themselves activity schools.

When we went to some of the schools that had been thought of as controls, we found a range too. It is true that more of the classrooms in the activity schools showed the characteristics of the original plan, but we also found some of these characteristics in the so-called control schools. If we had taken on face value the notion that an activity school actually operates according to the original design and had tried to make further appraisals, we would have drawn the wrong inferences. We based our study of the activity schools on three groups of classrooms—rooms that had around 55 out of 61 characteristics, rooms that had within 30 to 55, and rooms that had less than 30. We compared the kinds of learning being attained by these three groups. This procedure illustrates the examination stage, at which we check to see whether or not the curriculum plan is in operation before taking the more expensive step of evaluating pupil achievement in a particular program.

Goodlad and his group illustrated this procedure in the report on new curriculum materials and programs which they made for the Ford Foundation. He and his students visited classrooms to determine whether or not the programs under consideration were actually in operation. Take the Physical Science Study Committee (PSSC) as another example. Zacharias and his group have made a very explicit set of statements about the purposes of PSSC, including the notion that they have of what physical science learning is, the emphasis upon inquiry, the emphasis upon deriving generalizations from laboratory experiences, and the effort to use these ideas in interpreting other phenomena.

When Goodlad and his students talked to the teachers in a sample of classrooms they found that some understood or largely shared the views represented by the PSSC group and were operating the PSSC in that way, because their objectives were similar. At the other extreme they found teachers who viewed PSSC materials as just another textbook to be read and memorized. It was necessary to distinguish between those classrooms in which the purpose,

methods, and materials were consistent with the PSSC program and those which started out with the idea that anything that got into the hands of pupils was to be memorized and repeated back, because the programs which resulted would reflect the difference in approach.

In connection with this stage of evaluation, one should interview a sample of students to see how they conceive the objectives of their course (as indicated by the assignments they do and what they are trying to get out of the course) and to see the extent to which the learning is already operating. The next step is to find the extent to which this learning is actually taking place. In the case of PSSC, how far can these young people carry on inquiry learning? Can they use the equipment of the laboratory and simple apparatus to find out things for themselves? To what extent can they apply and understand the concepts taught in the course? To what extent can they apply these concepts in explaining particular phenomena?

These stages show the need for specialized tests for use in examining the effectiveness of particular procedures. The first stage of evaluation must determine whether or not the program is following the original plan. At this point the question is not a general one, such as, "Have they learned to read?" If, for example, a language program is developed on the stage-by-stage principles of modern structural linguistics, then the first question is whether or not the program is operating in the ordered sequence expected. Are the teacher and the pupils moving in this direction? Are they doing the tasks that are supposed to be part of the learning program? The next question at each stage is, are these learnings actually taking place? Precise assessment tools are required to probe into particular procedures or devices to see how they are working.

Assessing Learning Materials and Procedures

Another type of assessment is the appraisal of the actual effectiveness of various kinds of learning materials and procedures. The probing that takes place in the operations analysis to see how the material is actually being used and whether the step by step learning program is taking place has already been accomplished. Now we examine the degree of transfer, the extent to which students have learned the concepts they are now supposed to be generalizing and using in non-classroom situations.

It seems obvious that the whole purpose of education in the school is not to develop a person who can behave in desirable ways within the school but to develop a person who has acquired ways of thinking, feeling, and acting that are relevant to a wide range of human experience. What does he read outside of school? How well does his learning in the classroom serve him in the home, on the playground, in the community, or at work?

The function of the school's teaching is to develop young people whose behavior outside the classroom is effective and significant. Therefore, in appraising the relative effectiveness of curriculum materials or programs, one goes beyond a checking of program and purpose to consider whether the learnings are generalizable to life outside the school. The Progressive Education Association's Eight-Year Study, for example, followed a group of high school graduates into college and occupational roles to learn the extent to which they were able to utilize ways of thinking, feeling, and acting that the school had tried to develop.

We are all familiar with the general principle that any measures of education should be based upon educational objectives—what kind of learning are we seeking? Thirty-eight years ago, when Paul Diederich and I began some of these efforts in the Progressive Education Association, much was said about determining educational objectives. We talked about educational objectives at a level so general that such objectives represented desirable and attainable human outcomes. Now, as the people from conditioning have moved into an interest in learning in the schools, the notions of behavioral objectives have become much more specific.

As far as I know, one cannot very well teach a pigeon a general principle that he can then apply to a variety of situations. The objectives for persons coming out of the Skinnerian background tend to be highly specific ones. When I listen to Gagné, who is an intelligent and effective conditioner, talk about human learning objectives, I wince a good deal because he sets very specific ones. I know that we can attain levels of generalization of objectives that are higher than that.

As a graduate student at Chicago 42 years ago, I did a study with Judd, who was at that time arguing with Thorndike over the principle of transfer in learning. Thorndike had demonstrated that transfer was not automatic among the formal disciplines; a person could take a course in Latin and not be able to handle other kinds of languages any more effectively.

Thorndike reached the conclusion that every objective had to be very specific, like conditioning objectives. His first treatise on the psychology of arithmetic established some 3,000 objectives for elementary school arithmetic. Judd, however, had come out of the social psychology tradition, having studied with Wundt at Leipzig. His view was that generalization was not only possible but was essential in education. The task he assigned me was to check on Thorndike's view that the addition of every one of the 100 pairs of one-digit combinations had to be practiced by the learner before he could add all of the pairs. The design of my study was to take the principles of grouping for addition and help pupils see them. I noted that five and two, and six and one, and zero and seven, and three and four all total seven and had the students practice 21 out of the 100, emphasizing that each operation illustrated a general principle. I found that the youngsters in the experimental group who had practiced on only 21 illustrations did just as well on the average over the sample of the total 100 as the pupils who had practiced systematically every one of the 100.

The possibility of generalization is of course not new to the reader of this booklet. In curriculum development we now work on the principle that human beings can generalize, so they do not have to practice every specific. The question is at what level of generalization do we set up objectives. There are overgeneralizations you can immediately see; for example, the use of "you" for both singular and plural forms often confuses students in grammar exercises. The problem of the effective curriculum maker and teacher is to figure out the level of generalization that is possible with a certain child or a certain group, and then to establish objectives based on reaching that level of generalization. You will have 20 objectives perhaps, but not more. The conditioning view, based upon specific situations and practices, may involve several hundred objectives for a course because specific practices must be used to accomplish each aspect of the conditioning.

Discovering Problems

Assessment to discover problems before it is too late to deal with them is important to any good school or school system. This type of assessment requires not only the devices used in appraising relative effectiveness and exercises that assess transfer but also the development of baseline data. Baseline data enable us to talk

about change in terms of learning progress from grade to grade, rather than merely to assess the student's learning at a single given point.

The usual achievement tests have been based on such a wide range that they do not focus sufficiently. They do not present large enough samples of items at each grade or achievement level to determine whether or not learning is taking place. The lack of effective assessment tools has been one of the great problems in the effort to appraise Title I programs. The wide-range achievement test which has a perfectly useful purpose in getting means and relative standing and in identifying individual differences concentrates its exercises toward the middle level of difficulty. Only about five percent of the exercises on a typical achievement test we examined fell in the range of the lowest quarter of the age group for which they were intended. Five percent is too small a sample of exercises to find out whether that group is progressing, say, from the first grade to the second grade, and moving from year nine to year ten to year eleven.

If you are going to monitor a program, you need to develop exercises that are appropriate for each of the sectors of the age group that is being worked with. How does it happen that we cannot find enough of these exercises at the lower level? How does it happen that there has not been more emphasis upon this problem? We discovered that as long as the concern is with relative standing and means, the tendency of test constructors is to get more easy items simply by putting more cues in the written instructions. For example, a student can readily answer the question, "What color is blue vitriol?" without knowing a thing about copper sulfate.

The first round of the exercises being developed by the contractors for our national assessment of educational progress had an insufficient number of exercises designed to indicate the learning level of the lowest third of the age group. We had to bring in people who had been working with Head Start and other programs with disadvantaged children and knew how to communicate with them because we discovered that the children did not know what the exercises were. Furthermore, many nine-year-olds said they do not read well enough to understand the tests, so we put the exercises on audio tapes to be played as the youngsters read the test. We do not want the difficulty of an exercise to be the difficulty of reading unless it is a reading exercise. We found that many lower income

youngsters can really do computation because they have sold papers and done other things; yet, they might score zero on a written test because they did not understand the language of the directions. We have had to develop new exercises that are appropriate for the lower class sectors of the population.

If we are going to appraise effectiveness of programs, we have to focus on the sample of behavior that we want information about. We need specialized test samples wherever we have specialized problems.

College Admission and Placement

Assessment is important in college admission and placement. If we have a big pool of people graduating from high school with only a few going on to college, we can talk about skimming off the cream of the crop and then talk about single dimensions; but the situation now is that the prestige institutions could select their entire enrollment from the top five percent, taking the scholastic aptitude as the sole measure. However, colleges which select only from the top five percent report that they get a relatively uninteresting group of students, the students who have worked primarily to make high scores on scholastic aptitude tests.

For the nation at large, the demands for manpower and the opportunities for a higher education are so great that the problem is not that of selecting a small number. Fifty percent of the high school graduates are now going on to some post-high school education. The problem is to identify for them and for post-high school institutions what kinds of talents and interests they have. Such identification demands a much broader range of assessments than the single dimension of percentile rank. Holland finds, for example, that information about the things they have done—I've driven a car; I've led meetings; I've been an officer—are highly indicative of talents that can be developed further. There is a whole range of possibilities in admission and placement criteria for post-high school institutions that we have not utilized. Some people conceptualize the problem as the identification and selection of a narrow range of talent to go to particular colleges; those of us in the large state universities and in the junior colleges have a very different problem. If we are going to be helpful to the students, we need to identify a wider range of talents and to help our students to develop these talents.

Furthermore, the danger of admission and placement on the basis of prediction from scores is that you are predicting the students who can get along in the institution without having to make any effort; neither the student nor the institution has to do anything. This is one great danger of social institutions. The first generation of a social institution is one devoted to developing programs that will help the class which established it. The subsequent generations worship the program and see that they get only clients who fit into the program. A wide range of human beings need to be educated. We want colleges which make an effort to effect change, not colleges where the students get high grades without changing their ways. The problem is not predicting how to get along in a static college but how to get persons who can help to work the system for its own improvement. The students and the college can change and develop for greater effectiveness.

We need to identify the many purposes of assessment: to serve the teacher in the day by day work of the school, to serve supervisors and curriculum personnel developing and monitoring programs and materials, to serve the youngster in clarifying his own plans and programs, and others. No single kind of test or device is helpful for all of these purposes.

Problems in Assessment

What are some of the concepts which interfere with the development of effective assessment? The mean relative standing of individuals, or of individual schools (that is, the score in terms of "I am at the ninetieth percentile," or "this school is at the seventy-fifth percentile"), has very limited usefulness; a test designed to determine a mean tends to concentrate exercises at the middle level in order to get an accurate mean rather than to try to serve some of these other purposes. A second inhibiting concept is the notion of scaling in difficulty, based on the concept of learning as a normal growth process.

This is not the way we teach. We teach with the notion that there is something to be learned; it is the understanding of a concept, it is the development of a particular skill, or it is the acquiring of a new insight into human relations. We need to discover whether or not the youngster is acquiring this skill or this understanding or this attitude; it is not a scaling problem but a problem of sampling well enough the basic notions of a concept or the basic aspects of a skill to see whether the learning is taking place.

Prediction in most cases is based upon a static criterion. In our present stage of development, we recognize the dynamism. It is our role as educators to get greater dynamism, so that schools and colleges will move away from their past reliance on static predictions. We are committed too often to the notion of a test as something that takes time away from learning, so we have concentrated upon relatively short omnibus tests which do not sample anything well enough to serve most of the purposes that we have talked about. We have thought of testing every pupil; but for many of the purposes that I have outlined here, it is quite possible to test and interview a sample of six youngsters chosen from a class of thirty to represent different levels of background or different sorts of abilities. By interviewing some students, we could find out more about the learning that is taking place and the attitudes which are developing than we could by giving 30-minute test exercises to every pupil for this purpose.

The idea that everything that we find out through testing or sampling is to be fitted into a grade or award or punishment for the student stands in the way of meaningful assessment. Information should help us, rather than allow us to say, "We predict this for the youngster," and "Well, you made a B when you should have made an A."

We have limited ourselves too much to paper and pencil tests because we have wanted a test which could be administered to all of the students at one time. We are finding in the national assessment, where we can take samples, that it is possible to have youngsters respond orally, as Professor Diederich and his group have shown. We have youngsters listening and responding to music and recording their own musical efforts. We do not have to limit ourselves to prediction.

We have been limited by the overuse of selection-type tests such as multiple response, true-false, or some other kind of selection. While selection-type tests are very useful for some purposes, we can get a variety of information in other ways. The youngsters' own statements can be assessed in terms appropriate for the purpose. When we begin our assessment with a question—"What information do we need about students and about the learning situation for each of these purposes?"—it seems to me that we are on the road to improving our assessment.

Language, Rationality, and Assessment

ROBERT E. STAKE

IT IS my belief that we are not very effective in assessment because we are not very effective at formal communication. If I make such a claim and then support it (in *my* communication with you) with fastidious reasoning; if I cite just the right number of illustrations and speak with clarity and persuasion—then I weaken the claim. To support my claim that language is our shortcoming I will commit certain ambiguities of expression, I will violate some conventional definitions of terms, and I will read these unending passages in a mesmerizing drone. If I am successful, you will emerge from our afternoon session unable to recall a single confrontation with Truth. You will be convinced—as you have been at many a conference—that “Words Forever Fail Us”; which, of course, is my point.

Several years ago I was dismayed by the consternation shown by some experts in our field about the distinction between measurement and evaluation. As far as I was concerned the differentiation was an instance of “nit-picking.” I said, “Measurement implies evaluation. Testing just is not testing unless there is test interpretation. No ‘assessment’ occurs without an underlying intent to generalize.” I have joined the “nit-pickers.” Now I rally to the distinction. I want us to think of “something more” when we think of evaluation. I want us to think about the desirability of a student’s response as well as the quality of a student’s response.

And there is a second distinction. Most of my colleagues think of evaluation as measurement of individual student progress, but I want to focus some evaluation on individual school progress, and some on individual nation progress. I think it is important to define

evaluation differently than would most measurement specialists. My hortatory working definition goes like this:

As evaluators we should make a record of all of the following: what the author or teacher or school board intends to do, what is provided in the way of an environment, the transactions between teacher and learner, the student progress, the side effects, and last and most important, the merit and shortcoming seen by persons from divergent viewpoints.

I see a useful distinction between measurement and evaluation. Am I able to make a useful distinction between measurement and assessment? I like to think of assessment as one form of measurement. Going along with Jum Nunnally (1959), I say that assessment is direct measurement, in contrast to psychometric testing, which almost always is indirect measurement. Assessment, as represented by the National Assessment Project (Tyler, 1965), pertains to direct measurement of performance on important reference tasks. Both psychometric testing and assessment are useful techniques for gathering information.

Curriculum Evaluation

Here I am going to talk about something broader. I will discuss inquiries into the worth of any instructional program. Such inquiries depend on direct assessment, on objective testing, and on subjective judgments. I will call such an inquiry: evaluation. If what I call "evaluation" is much different from what you call "assessment" then perhaps I should retitile this paper: "Language, Rationality, and What I Call Evaluation."

Ralph Tyler has done a magnificent job of describing the multiplicity of evaluation roles. One of the distinctions most helpful for understanding a theory of evaluation, I believe, is the distinction Mike Scriven (1967) makes between the *roles* and the *goal* of evaluation. The goal of evaluation is always the same: to determine the worth of something. The roles depend on *what* that something is and on *whose* standards of value will apply. A student's performance can be evaluated by those considering his admissibility to advanced training. That is one role for evaluation. A million student performances can be evaluated by persons concerned about a nation's academic curricula. Competing textbooks can be evaluated—that is, their relative merits can be examined. Environments

PROGRAM RATIONALE	DATA FOR THE EVALUATION OF AN EDUCATIONAL PROGRAM			
	Intents Sources	Observations Sources	Standards Sources	Judgments Sources
ANTECEDENTS Student Characteristics Teacher Characteristics Curricular Content Curricular Context Instructional Materials Physical Plant School Organization Community Context	A			
TRANSACTIONS Communication Flow Time Allocation Sequence of Events Reinforcement Schedule Social Climate				D
OUTCOMES Student Achievement Student Attitudes Student Motor Skills Effects on Teachers Institutional Effects		B	C	
Example A: Manufacturer Specification of an Instructional Materials Kit Example B: Teacher Description of Student Understanding Example C: Expert Opinion on Cognitive Skill Needed for a Class of Problems Example D: Administrator Judgment of Feasibility of a Field Trip Arrangement				

FIGURE 1. Illustration of Data Possibly Representative of the Contents of Four Cells of the Matrices for a Given Educational Program*

*Adapted from: Robert E. Stake. "The Countenance of Educational Evaluation." *Teachers College Record* 68 (7): 529; April 1967. Used by permission.

can be evaluated. Educational goals can be evaluated. By mentioning these I illustrate the roles that evaluation can play.

By this definition it is inappropriate to claim that all evaluation should focus on student performance. It is inappropriate to claim that all educational evaluation should focus on goals specified by the curriculum designer. There are other important roles for evaluation than to determine the extent to which teaching objectives have been attained. People who set objectives—programmers, teachers, experimenters—may be particularly interested in attainment of the goals *they* specified; but others have other goals. A group of taxpayers, philosophers, or students will choose to look at different criteria of merit, and will have different standards against which to make value judgments. As people have different uses for evaluation information, the roles of evaluation will differ.

Most people will use evaluation information, directly or indirectly, for making decisions. Curriculum developers make decisions; teachers make other decisions; counselors make still other decisions. If we can anticipate the choices, we will have some of the guidelines for an evaluation plan. Daniel Stufflebeam (1966) discusses evaluation for decision making, so I will not; but I do want to talk about the generalizability of evaluation information. Let me summarize what I have said about evaluation so far by suggesting three questions that should be asked prior to drawing up an evaluation plan:

1. What is the entity that is to be evaluated?
2. Whose standards will be used as reference marks?
3. What subsequent decisions can be anticipated?

Answers to these three questions should be sought prior to planning the evaluation.

Now let us inventory the data I believe the evaluation plan should call for. In my paper, "The Countenance of Educational Evaluation" (1967a), I suggested use of a huge matrix of evaluation information. A representation of this matrix is included in Figure 1. You will see there an array of *row entries* that help identify the many characteristics of the instructional program to be evaluated. The evaluator must choose the variables to be described and judged.

The column entries in this matrix identify separate *sources* of information: teachers, administrators, counselors, professors,

parents, and so on. My matrix does not say which sources and which variables are important. It just reminds the evaluator that he has to pick and choose among a potential deluge of information. Obviously, information to fill the thousands of sub-cells of this matrix could not be obtained in any one evaluation study. A principal task of the evaluator is to concentrate attention on variables that are related to the goals of his audience, variables leading to decisions, and variables that are available—within his budget—from appropriate sources. (I might add that evaluators will have different degrees of interest and talent for measuring different variables. I think the sponsor of an evaluation study should pay considerable attention to what it is that the evaluator *likes* to measure.)

If a set of instructional materials is to be evaluated, the variables might be organized as shown in Figure 1a. Here, great attention is paid to the "conditions of use" for the textbook or science kit or whatever entity is being evaluated (Stake, 1967b).

Back in the grid in Figure 1 we have 12 major cells—plus a thirteenth in which to represent the rationale. We find out what different goals people have; I call these *intents*. We note our perceptions of what actually happens; I call these *observations*. We list statements by certain experts as to what should happen in a situation like ours; these are *standards*. And we gather data on how people feel about aspects of our situation; and these are our *judgments*.

In any curriculum—even in the briefest lesson—different people have different intents. And there are many relevant observations that we can schedule. There are many standards that could be useful to the audience that will receive our report, and there are many judgments that will be made. These are the classes of descriptive and judgmental data that I believe are needed in curriculum evaluation.

Congruence and Contingency

I perceive two principal ways of processing the descriptive evaluation data: finding the contingencies among antecedents, transactions, and outcomes and finding the congruence between Intents and Observations. The processing of judgmental data follows a different model. The first two main columns of the data

matrix in Figure 1 contain the descriptive data. The format for processing these data is represented in Figure 2 (Stake, 1967a).

Subdivision	Examples of variables in the subdivison
Conditions of Use	
Local Circumstances	
Student types.....	(background, aptitude, aspiration . . .)
Teacher type.....	(experience, style, personality . . .)
Type of school.....	(physical plant, intellectual climate . . .)
Type of community...	(support of schools, attitudes, controersy . . .)
Curricular Context	
Subject matter	
coverage.....	(concepts, structure, methods of inquiry . . .)
Instructional aids	
available.....	(library, models, maps, equipment . . .)
Concurrent	
course work.....	(sequence and time allotment, projets . . .)
Classroom Transactions	
Teaching strategies...	(discourse, inquiry, assignments . . .)
Student-teacher	
interaction.....	(information, flow, counseling . . .)
Student-student	
interaction.....	(social climate, reaction to authority . . .)
Incentives,	
grades, etc.....	(motivation, goal orientation, testing . . .)
Results of Use	
Gain in Student Competence	
Knowledge.....	(data, understanding, application . . .)
Skill.....	(problem solving, communication . . .)
Incidental learning...	(synthesis, learning sets, side effects . . .)
Change in Student Attitude	
Interest.....	(opinion, avocation, exploration . . .)
Commitments.....	(prejudice, aspiration, advocacy . . .)
Effects on Staff	
Teacher changes.....	(insights, revision, grievances . . .)
Administrative	
changes.....	(organizational rearrangements pernitted...)
Other Effects	
Institutional effects...	(prestige, solidarity . . .)
Community effects....	(controversy, dedication, esprit . . .)

FIGURE 1a. Subdivisions of Information Classes for Evaluating Educational Products*

*From: Robert E. Stake. "A Research Rationale for EPIE." *The EPIE Forum* 1:7-15; September 1967.

Descriptive data

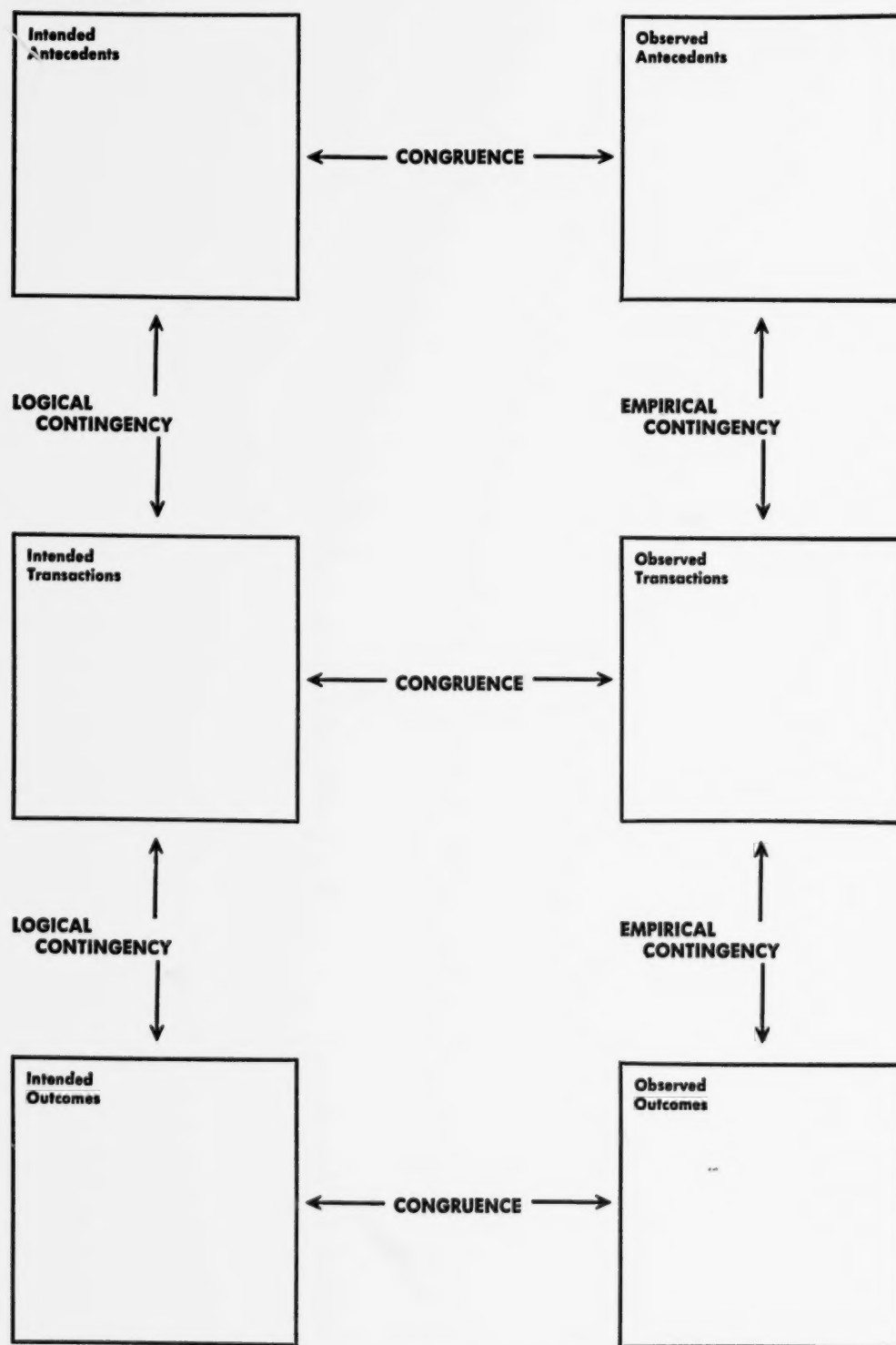


FIGURE 2. A Representation of the Processing of Descriptive Data*

* Robert E. Stake. "The Countenance of Educational Evaluation." *Teachers College Record* 68 (7): 533; April 1967. Used by permission.

Intents and Observations are *congruent* if what was intended actually happens. To be fully congruent the intended antecedents, transactions, and outcomes must be identical with the observed antecedents, transactions, and outcomes. (This seldom happens and often should not.)

Some evaluation studies concentrate only on the congruence between intended and observed outcomes. If our purpose is to continue a good curriculum or revise a poor one we should know about congruence of antecedents and transactions as well. Working horizontally in the data matrix, the evaluator will compare the information labeled Intents with the information labeled Observations—he will note the discrepancies and describe the amount of congruence for that row. Congruence does not indicate that outcomes are reliable or valid, but that what was intended did in fact occur.

It should be obvious that congruence and lack of congruence are more easily discovered when the same language is used to describe the goals and the actual operations. One way to synchronize language is to focus on teacher, administrator, and student behaviors.

So much for the moment for congruence. How about contingency? Contingencies are relationships among the variables. An evaluator's search for contingency is in effect the search for causal relationships. These are what Hastings (1966) called the "whys of the outcomes." Knowledge of what causes what obviously facilitates the improvement of instruction. One of the evaluator's tasks is identifying outcomes that are contingent upon particular antecedent conditions and particular instructional transactions.

For as long as there has been schooling, curriculum planning has rested upon faith in certain contingencies. Day to day, every teacher arranges his presentation and the learning environment in a way that—according to his logic—leads to the attainment of his instructional goals. His contingencies, in the main, are logical, intuitive, supported by a history of satisfactions and endorsements. To various degrees teachers test out these contingencies. (Some of us would have them use more deliberate, more standardized, confirmation techniques.) Even the master teacher and certainly less experienced teachers need to examine the logical and empirical bases for their "believed-in" contingencies. Do colleagues agree that their plans are logical? Have experts found such arrangements and teaching methods to "pay off" in that way?

One first step in evaluation is to record the potential contingency. A film on floodwaters may be scheduled (intended transaction) to expose students to background for understanding conservation legislation (intended outcome). Of those who know both subject matter and pedagogy, we ask, "Is there a logical connection between this event and this purpose?" If so, a logical contingency exists between these two Intents.

Whenever Intents are evaluated, the contingency criterion is one of logic. To test the logic of an educational contingency, evaluators rely on previous experience, perhaps on research experience, with similar observables. No immediate observation of these variables, however, is necessary to test the strength of the contingencies among Intents.

Evaluation of Observation contingencies depends on empirical evidence. To say, "This arithmetic class progressed rapidly because the teacher was somewhat but not too sophisticated in mathematics" demands empirical data, either from within the evaluation or from the research literature. The usual evaluation of a single program will not alone provide the data necessary for contingency statements. Relationships require variation in the independent variables. What happens with various teaching treatments? Here, too, as Ausubel has contended (1966), previous experience with this content and with these teaching methods is a basic qualification of the evaluator.

The contingencies and congruences identified by evaluators should be judged as to importance by experts and interested parties just as the descriptive data are. The importance of non-congruence will vary with different viewpoints. The school superintendent and the school counselor may disagree as to the importance of cancellation of the scheduled lessons on sex hygiene in the health class. Here is an example of judging contingencies: the degree to which teacher morale is contingent on the length of the school day may be deemed cause enough to abandon an early morning class by one judge and not by another. *Perceptions* of the importance of congruence and contingency deserve the careful study of the evaluator.

We could now shift over to the right-hand side of the grid and consider the processing of standards and judgments for evaluation purposes. I am not going to do that here, for several reasons, one of which is that I really do not know much about processing judgments. I discussed this briefly in my "Countenance" paper, but I am sure I gave the reader little guidance for that important step

between reading the evaluation report and making the educational decision.

Generalizability of Findings

Looking back at the emphasis I have given so far to rationality and to specification and to contingencies, the reader may be thinking that I cannot see the distinction between instructional research and evaluation. I do have difficulty drawing a line separating them. In fact, I am now going to draw a line connecting them. I see inquiry about instruction placed on a continuum. At one end, the findings are quite generalizable. At the other end, the findings are less generalizable. The line is a continuum of generalizability. I will put four important points on this continuum, one for instructional research, one for formative evaluation, the other two for summative evaluation and institutional evaluation.

Instructional Research	Formative Evaluation	Summative Evaluation	Institutional Evaluation
Generalizability of Findings			

All four of these kinds of studies can be called applied research. All of them seek important information for the conduct of education. No one of them is necessarily more abstract than the other—they all deal with the concrete, the practical, the everyday components of education. They do differ with regard to generalizability. Findings from *instructional research* are more generalizable than findings from the others. Classroom studies of problem solving, modeling behavior, achievement need, content sequencing, and reinforcement are usually instructional research studies. These studies are expected to generalize—extensively if not completely—over subject matters, over school settings, over student types, over teacher types, and over time.

Formative evaluation leads to findings less generalizable than instructional research, but more generalizable than summative evaluation. Formative evaluation seeks information for the development of a curriculum or instructional device. The developer wants to find out what arrangement or what amount of something to use. Western Electric does formative evaluation when it tries out various plastics to determine which will make the more durable casing for

a particular telephone. The Educational Testing Service does formative evaluation when it researches the effect of item vocabulary-difficulty on the discriminability of the National Teachers Examination. The BSCS Biology Project does formative evaluation when it studies the number of positive and negative instances of "genetic mutation" needed to get the concept across to an anticipated student body. The developer assumes that for subsequent revisions of that device the findings will hold. He assumes—sometimes incorrectly—that the findings are not specific to student types and teacher types; certainly he is comfortable in the belief that a separate finding is not necessary for each teacher and each student. Findings from the formative evaluation study are generalized over school setting, teacher and student types, and within the various versions of the particular instructional device, but are not generalized across subject matters and curricula.

Now, what about *summative evaluation*? I see summative evaluation aimed at giving answers to an educator about the merits and shortcomings of a particular curriculum or a specific set of instructional materials. This decision maker has little opportunity to modify or revise those packages. Particularly in the future, as computers, audio-visual equipment, laboratory kits, and other distantly-developed units are increasingly used, the local consumer will be in no position to modify or revise them. He must learn about them as they are.

The monthly *Consumer Reports* does a good job of summatively evaluating refrigerators and various kinds of cameras. It expects these products to be used in a variety of ways, but each product will be accepted as a unit. Little rearrangement of its components is to be expected. *Buros' Mental Measurements Yearbook* does a good job of summative evaluation of standardized achievement tests. The findings are expected to be generalizable across large numbers of schools, teachers, and students. A large responsibility for the local educator remains in determining how similar are his uses, how similar are his teachers and students, to those described in the summative evaluation report. Particularly in this matter of deciding the appropriate bounds of generalization, words have often failed us.

Institutional evaluation, like summative evaluation, is aimed at a specific curriculum or instructional device but, in addition, is oriented to a specific setting—with its distinctive goals, classrooms, teaching staff, and student body. The evaluation of this Work Con-

ference would be an example of institutional evaluation. The evaluation of the Tampa, Florida, First Grade Reading Program would be. So would the evaluation of the Peace Corps. For any institutional evaluation the curriculum, setting, staff, and students are specified. They may not all be examined but they are fixed. External norms are not of highest importance. The reader of the institutional evaluation report has relatively little need to know how his educational setting and personnel compare to others. He is little concerned about generalization to other settings or to other curricula. He is concerned about congruences and contingencies for inputs and outcomes for a specified teaching situation.

All four of these kinds of studies can be based in the curriculum project or in the local school. What I have in mind is not so much the classification of different evaluation efforts but the importance of generalizability. A primary consideration in organizing an evaluation study is deciding on the degree to which the findings should be generalizable across curricula, school settings, teachers, and students. Different limits, of course, call for different data-gathering plans.

One of the reasons that many teachers and administrators pay little attention to research and evaluation studies is that they do not believe the findings will be generalizable to their situations. There is a common belief among educators that ideal programs can only be those tailored to the local community, to a particular teaching style, and sometimes, to each and every child separately. If this belief is well founded, all methods of educational evaluation are going to be very expensive. If instructional devices and situational features do yield an interaction effect, our studies become much more complex. If there is little generalizability—if, for example, the worth of a curriculum is highly dependent upon the value commitments of the teacher—then the value commitments of the individual teachers must be studied.

Educators do expect interactions such as these; yet apparently there is very little research evidence to substantiate them. Of course teachers are different. There is evidence that each should be paced individually. Different step sizes are appropriate for different teachers and different students. Yet this is not to say there is a disordinal (cross-over) interaction between program and personnel. This does not say that one version of the new mathematics program will be better with one population and another program better with a second population. At the present time there is little

evidence that age, race, or sex and little evidence that cognitive style or temperament are keys to selecting better instructional treatments. We need to know more about these bases for individualization of instruction before we will know how customized evaluation plans will have to be.

Rationalism and Empiricism

Now let us look at some similarities and differences between research and evaluation methods. Scientists are observers. Evaluators are observers.¹ Both are seeking generalizations. A majority of scientists are manipulators. Developers of educational devices are manipulators. Evaluators are not. They are only free to manipulate themselves into better positions for observation. Even so, their presence may affect the outcomes. Scientists and evaluators alike must worry about such reactive effects.

Scientists, at least basic scientists, are not burdened by the demands of social utility. Their responsibility is revelation. A major scientific contribution reveals things in a different light, from a different perspective, in a different language—almost always in a way more parsimonious than the last. The evaluator need little concern himself with parsimony—in fact, he should err on the side of complexity and redundancy and detail. The educational evaluator's obligation is not to discover the essence of human learning, but to discover the diversity of viewpoints and explanations of what is going on in the school. His obligation is not to find the simplest explanations but the ones most amenable to control. The evaluator does shoulder the burden of utility. He must anticipate the uses of the evaluation.

Both the scientist and the evaluator hold "rationality" in high esteem. The inquiry methods of both the scientist and the evaluator are orderly, constrained, deliberate, based on reason. The more respectable designs—among scientists and evaluators—are those that proceed from theory to hypothesis testing, from general rule to specific instance, and from program rationale to specific practice. It is said that an evaluation plan should be rational.

Now the important contradistinction here is not between rationality and irrationality. (To most people *that* distinction is

¹ Of course this does not say that the same man cannot be both scientist and evaluator. Even in the same study, each of us may choose to play both roles.

parallel to the distinction between good and bad.) The important distinction is between rationalism and empiricism. Here are two systems for seeking generalizations. According to the rationalist, we should think first, plan ahead, anticipate the generalization, draw a flow chart or map or blueprint—then act, then test out the idea, confirm the generalization. According to the empiricist, we should act first, observe, build up a backlog of experience—then abstract, then infer, the generalization. The rationalist would have us invest more in planning; the empiricist would have us invest more in experience.

The distinction I am making might be more familiar to the reader as the distinction between the hypothetico-deductive method of inquiry and the heuristic method of inquiry. I find it easier to use the terms rational and empirical.

For evaluation, as for science, these alternate methods can be compared as to usefulness. We can compare the results of inquiries based upon a more rational approach with results of inquiries based upon a more empirical approach. Which do we like better? It is reasonable to anticipate that one will be narrower in coverage than the other. Which one? It is reasonable to anticipate that one will have more internal consistency than the other. Which one? In both cases, choice (a). The rational study, with its greater focus, is always in danger of ignoring important accoutrements. The empirical study, with its greater scope, is always in danger of having relevancies obscured by the irrelevant. Both can be powerful methods of inquiry; both can be carried to unwarranted extremes.

Although it may not be possible to have too much experience, it is possible to emphasize "knowing the classroom" so much that no attention is given to putting experience into order. An evaluation plan may fail because it deals only with vague, personal impressions.

Although it may not be possible to be too reasonable, it is possible to emphasize rationality so much that encounters with reality are unduly delayed or narrowly conceived. An evaluation plan may fail because it squanders its resources on organization, on instrument development, and on delimitation of the problem.

The technologist, the measurement expert, the proponent of rationalism says, "You misunderstand us, sir. Give rationalism a chance to work before you conclude it will not." No true empiricist could refuse that plea. As for me, personally, I do not want to be counted among those who are sure the worst will happen, who

expect the misuse of technology to outrank the gain. Let us think about what rationality may do for us, good and bad. Maybe we should give it a "real" try. Like the empiricist says, "The making of something better depends on trying out things that we don't know much about." What does the contemporary rationalist want us to try out first?

The cornerstone of contemporary rationalism in education is the statement of objectives. Most plans call for the formulation of objectives prior to the operationalization of them. Most plans involving students call for the statement of objectives in behavioral language.² The argument is made that teachers do not analyze their teaching; that they are not aware of many of the goals that they really are teaching for. It is claimed that teachers should commit themselves to "modification of specific behaviors." Unfortunately, it is not at all a demonstrated fact that teachers teach better when they state their objectives behaviorally and when they critically analyze their own teaching behavior.

When I look at teachers who seem to be doing a superb job of teaching my advisees and my children, I seldom find evidence that they are conceptualizing their task in behavior-modification language. Many seem little disposed to analysis of the teaching act. I do not know of any studies of preservice or in-service teacher education that suggest that teaching effectiveness is increased by allocating more of the teacher's time to planning and analytical evaluation. There is a lack of congruence between the expectation of the behaviorists and my observation of classroom teaching. It has prompted me to write an informal position statement. This statement summarizes a number of points I am attempting to make here.

Educational Objectives: A Position Statement

1. A great number of educational objectives are simultaneously pursued. The high-priority, immediate objectives should usually be apparent to teacher and learner alike. Occasionally, either will do better

² Behavior is associated with overt personal experience, so behaviorism traditionally has been associated with empiricism. Goal statements of behavior, however, are often outside the language repertoire of the educators involved, thus little associated with personal experience. An emphasis on any abstraction, e.g., a statement, is more characteristic of the rational than the empirical point of view. Thus, working with behavioral objectives is for most educators consorting with rationalism, not empiricism.

without being aware of them. High-quality education does often occur with educators having only an approximate realization of the objectives. Sometimes it will increase teaching-learning effectiveness to make participants more aware of objectives; sometimes it will not.

2. With all who share the responsibility of educating, there lies the responsibility for stating objectives, arranging environments, providing stimulation, evoking responses, and evaluating those responses. Yet each author and teacher does not share equally in those responsibilities. Time and talent are not available in limitless abundance to anyone. Each educator's assignment should capitalize on what he can do best. Few classroom teachers are skilled in stating objectives. Most are more highly skilled in adapting teaching to immediate circumstances, motivating students, and appraising responses. In the interest of effectiveness, seldom should they be required to formulate or conform to behavioral specifications.

3. There are more objectives to pursue than we can follow. Time and resources restrict us. We assign priorities to our goals in a highly informal way. This priority list is not the only determinant of the daily lesson or the minute-by-minute dialogue. Some moments are ripe for teaching toward an unplanned objective. A sound educational system is one which provides for occasional reassignment of immediate objectives to take advantage of the special opportunities that occur.

4. The development of a new curricular program or set of instructional materials often proceeds better by successive approximations than by linear programming. With successive approximations, major attention is given to getting an enterprise in operation, even though the initial runs are crude and faulty, so that corrections can be based on experience. With linear programming, major attention is given to planning, precise specification, and symbolic representation so that corrections can be based on logical analysis. Advice on curriculum planning should be oriented to the experiential and logical skills already developed in the developers or that can be readily obtained by them.

5. For creating lists of objectives, the technology of education should have some methods that rely on behavioral specification and symbolic delimitation and other methods that rely on illustrative examples and inferable definitions. We need methods by which educators and others can endorse, reject, or revise statements of objectives. Two colossal problems lie before us: how to *translate* global objectives into specific behavioral objectives and how to *derive* appropriate teaching tactics.

6. Our curriculum development projects and our evaluation studies seldom reach a satisfactory specification by asking educators to state their objectives. Educators' global objectives give little guidance to

teaching and evaluation. Their specific objectives ignore vast concerns that they have. In our present state the derivation of the specific from the general is some form of intuitive magic. Luckily this process often works pretty well. We need to understand it, to simulate it, not necessarily to replace it.

A Second Test

Let us go back to what the contemporary rationalist wants tried out in the classroom. He says teachers should stick to the lesson plan. Should a teacher be denounced because he does not stick to the syllabus? Departure from prescribed goals would be a sin indeed if we could just barely accomplish all our goals in the time available. The fact is, of course, that we cannot accomplish nearly all our goals. Furthermore, there are many important goals which can be pursued only when the situation is right, and for which it is difficult to create that situation. It is difficult to program many objectives, especially in the affective domain. Yet there are times when the classroom situation seems just right for teaching them.

Consider a teacher in an advanced biology class. A dialogue approximately like this occurred recently at the University of Illinois University High School. Miss Betty K. was teaching a small group of students about metabolism.

"... DNA (coded instructions) received from the previous generation are transcribed into RNA (again coded instructions) which ultimately are translated into specific molecules. What is unique about this whole system is the fact that each individual gets a unique set of coded instructions and ultimately ends up again with a unique set of proteins. Okay? Now with this understood, we can look at the details of metabolism, how we get ...

"Uh, wait a minute, I'd like to know how you can transplant, if each thing is unique, if each set of proteins is unique, how can you transplant an organ from one race to another. Like, for instance, in the recent heart transplants. They used a mulatto.

"Oh. Well ... do you remember what they did as they reported this case in the paper, before they prepped the person for the transplant?

"They lowered the antibodies, well, they lowered the resistance of the person to make antibodies.

"And what else?

"So he couldn't reject the heart.

"Okay, but what other tests do they perform on the donor before they would ...

"They had to have the same blood type.

"They ran tissue tests to see if the tissues were similar and same type of proteins.

"They also had to have the same size heart, so they wouldn't die because it was inadequate to pump the blood.

"I heard they can even transplant the organs from monkeys or something like that to human beings.

"Is something like that right?

"Uh, well, what do you mean by right?

"Well, I don't know, is it legal or moral, I mean you might be half-human and half-monkey by the time you're finished.

"I think it's, I don't know, it seems to me that if you're on your deathbed, then you're going to grasp at anything. If you can live for a little longer with a monkey heart, then it's probably best for you to use it.

"Well, these are the kinds of questions that you can't answer yes or no. Maybe we should pause and consider this kind of question because these are the kinds of . . .

"Like do you think it's right to have a monkey heart?

"Well, not that specific example, but this kind of thing. We should make these kinds of decisions, and how should they go about making decisions like this. . . ."

The opening statements of this transaction are analyzed in Figure 3 on the next page.

I have guessed at what was going through the teacher's mind during that exchange. On the two sides of Figure 3 are represented some of what is stored in the teacher's memory. Her objectives are of two kinds: she plans to stimulate her students in various ways (the s's) and she plans to work with certain responses that will occur (the r's). In some way she appears to compare the responses she encounters in the classroom to those she wants to occur. When an unusual response occurs she examines its potential for leading to some long-range objectives, then seeks other ideas and responses which more nearly approximate that goal.

It seemed to me that this teacher tried to provide opportunity for reflection and reaction. Quite unlike the linear and branching programmed instruction we know, hers was the operant conditioning paradigm. The teacher seemed prepared to identify, reinforce, and shape many kinds of responses. It was as if she had an inventory of immediate objectives and an inventory of long-range objectives. She set things up so that immediate objectives were attended to until there arose the occasional opportunity to work on a hard-to-program objective.

		Transaction		
		... DNA (coded instructions) received from the previous generation are transcribed into RNA (again coded instructions) which ultimately are translated into specific molecules. What is unique about this whole system is the fact that each individual gets a unique set of coded instructions and ultimately ends up again with a unique set of proteins. Okay?	Rationale	
Curriculum Guide			Long-Range Objectives	
Lesson Plans				
Short-Range Objectives				
Content	Response		Content	Response
. . . S S S	r r r . .	<i>Here the teacher checks to see if anyone wants further clarification on the background.</i>	. . S S S	r r r . .
. . . S S S	r r r . .		. . S S S	r r r . .
. . . S S S	r r r . .		. . S S S	r r r . .
. . . S S S	r r r . .		. . S S S	r r r . .
. . . S S S	r r r . .	Now with this understood, we can look at the details of metabolism; how we get . . .	. . S S S	r r r . .
. . . S S S	r r r . .	Wait a minute. I'd like to know how you can transplant an organ from one race to another. Like, for instance, in the recent heart transplants. They used a mulatto.	. . S S S	r r r . .
. . . S S S	r r r . .	Oh. Well . . .	. . S S S	r r r . .
. . . S S S	r r r . .		. . S S S	r r r . .
. . . S S S	r r r . .	<i>Here the teacher considers whether this digression has any potential merit, whether or not this might lead to goals difficult to "teach for" in other contexts.</i>	. . S S S	r r r . .
. . . S S S	r r r . .		. . S S S	r r r . .
. . . S S S	r r r . .	. . . Well, do you remember what they did (as reported in the paper) before they prepped the person for the transplant?	. . S S S	r r r . .
. . . S S S	r r r . .		. . S S S	r r r . .
. . . S S S	r r r . .	<i>The teacher is stalling for time at the start, but by the time she has completed the question she has decided to pursue the topic, at least a little.</i>	. . S S S	r r r . .

FIGURE 3. A Representation of Classroom Instructional Transaction Emphasizing Continuous Teacher Evaluation of the Situation and Potential for Revision of Immediate Objectives and Priorities

The operant conditioning paradigm begins with a desired response, voluntary on the part of the learner. Many educationally desirable responses can be elicited just by asking for them or by "fishing" for them; many cannot. While in Miss K.'s class, I encountered another example of these unusual, unexpected, pregnant responses. A student said, "Isn't that the sort of relationship you can show with a graph?" It is very difficult to teach graphing as language in contrast to graphing as a form of (shall I say) penmanship. *There* was an opportunity. A skilled teacher will seize the opportunity to reconsider objectives. She will reassign priorities to goals on the spot. She may do this without being aware of the old or the new priorities. She may do it because "that's just the way you teach." But there is no rule that the teacher must recognize the operant paradigm, or call it that, for it to be effective.

This conscious or unconscious review of objectives seems to me to be the important purchase we make in assigning curriculum control to the teacher. There are many advantages to external programming, e.g., writing lessons in advance as the programmed instruction people do or as the well-organized lecturer does, but these advantages should be weighed against the advantages of assigning control to teachers who are sensitive to conditions optimally suited for the pursuit of elusive, long-range goals.

I would like to make a point here that should have been made previously. Teachers and evaluators need not have the same commitment to rationalism. It is not undesirable for us to have a high majority of teachers whose style is intuitive, spontaneous, and empirical; and at the same time, to have a high majority of evaluators who are programmatic, deliberate, and rational. The successful practice of evaluation should not depend on teachers' being able to anticipate their information needs or to formulate their goals behaviorally. It is one thing for us to advise our colleagues in evaluation to commit themselves to rationalism; it is quite another to contend that teachers and curriculum supervisors do likewise.

I have examined here some contrasts between being rational and being empirical, particularly as they affect evaluation and as they affect the giving of advice to fellow educators. I singled out two of the propositions of the rationalists in education and found objections to both. This is only a small sample so we should not conclude that all rationalist advice is objectionable, or that empiricist advice is better. We need to examine carefully more advice

of both kinds.³ In the final section of this paper I want to talk about language.

The Language Barrier

Each evaluation has its audiences. We evaluate in order to tell those audiences about an instructional program. The quality of the evaluation will not exceed the quality of its communication. It is my contention that the greatest constraint upon evaluation today is the low quality of the language of evaluation. Our concern for goals is adequate, but our ability to represent goals is inadequate. Our talent for measuring educational outcomes is admirable, but our ability to convey their meaning is disappointing. Our ability to select the variables that people want to know about is often satisfactory, but the concepts we use are misunderstood. We are capable of restricting the subjectivity of our observations, but we are less capable of translating those observations into a language the audience can share with us. Our audience seeks certain information but we often misinterpret the needs.

How can they tell us? How can we talk to them? How can we indicate, for example, that the students are now more ready to participate in a formal learning experience than they were at the outset? What words can we use when we think we see, for example, a resistance-to-change on the part of a teachers association? It is true that these are measurement questions, but better observation schedules, better attitude scales, alone will not suffice. We need to improve the language, to talk more coherently to people about education. Without losing what precision of measurement we have, without jargon but yet without stirring up all the connotation of our childhood vocabularies, we need to increase our capacity to share meaning with others.

It is not necessarily sensible to say that we will teach them *our* language, nor necessarily sensible for us to translate, to mold our ideas into their language. Both or perhaps neither. I do not know how languages are nurtured, but ours must grow.

³ One basis for evaluating rationalist advice against empiricist advice relates to Kuhn's (1962) classification of scientific behavior: prescience (discrete experience); natural history (organized experience); normal science (testing theories and application); and extraordinary science (the breakthrough in theory). If a branch of education is devoid of theory, it is in a natural-history state of science and its practitioners might best rely on orderly experience, empiricism. If it has a substantial formal theory, its practitioners might be unwise to rely heavily on personal though orderly experience, wiser to specify purpose and plan on rational grounds.

Let me give some examples of what I think we need in the way of new or improved language. First of all, we need concepts. We have the concepts of achievement, verbal ability, grade equivalents, vocational training. I have mentioned the concepts of congruence and contingency. Contingency (the idea, not the label) is common to discourse on the quality of schools, but the concept of congruence is not so common. Illustrative of other concepts that may be worth developing are the concepts of colinearity of classroom proceedings and homework assignments, relevance of instances used in concept formation learning, student concern for the learning difficulties of other students, and community responsiveness to changes in extracurriculars.

Many of the concepts become better understood as we use indices or models to represent them. Second, then, we need indicators of many aspects of school function. The National Assessment Project proposed new indicator items of student competence. Norm Kurland of New York State proposed a number of scholastic indicators. Project YARDSTICK in Cleveland is looking for an index of school efficiency. The difference between indicators and test scores or measurements, of course, is their acceptability as standard representations of important concepts. The I.Q. is an indicator of intelligence, though no longer acceptable as such to some. The Achievement Quotient, purportedly an indicator of over- and under-achievement, has not proved to be useful in most situations. Average daily attendance has.

There are those who protest the use of such indicators because they are not error-free, because they oversimplify. But all language is an approximation to the thought process. Most words and descriptions are simpler than the phenomena they represent. The only total safeguard against miscommunication is no communication. I understand that there are some economists who have been protesting 30 years against the Gross National Product as an indicator of national productivity. Yet that indicator is useful. If its meaning needs refining, additional indicators can be used. There will be times when there is a real danger that an indicator will be misused and the consequences will be costly, so the indicator should be abandoned before it has really been tried out. Usually not, however. The lay of the land is not better understood by decreasing the number of bench marks. We need more. How to develop indicators and other forms of evaluation language is a question I will postpone until the very conclusion of this paper.

As a third component of language, I think that we need a better way of delimiting objectives. As I have said, I feel that neither the behavioral specification of goals nor the global summary of goals represents what the schools are trying to do. Either may be a suitable point of departure for developing more accurate language. Neither is satisfactory now.

A truly representative list of educational goals will contain competing and even contradictory goals. Goals compete with each other. Each pursuit costs something and the total of our resources will always be less than the cost of pursuing all goals. We have to choose among our goals. We assign priorities to them. We may do this unconsciously but we do it.

Some goals will be contradictory. We seek incompatible outcomes. We try to teach faith *and* skepticism. We try to instill deep appreciation, and yet provoke aspiration for something better. We hope that any one teaching effort will aid persons with different aims. We seek to serve a pluralistic society. Contradictory goals are to be expected in a pluralistic society. We cannot hope to pursue only goals that are perfectly complementary and universally wanted.

Evaluators should be alert to the fact that goals are changing. Our world changes. Our needs change. Our values change. Some of our goals change even as a function of what happens during instruction. A program evaluation is incomplete if it goes no further than designating several specific goals at time zero. To understand what is happening in a training activity and to ascertain its value, we are obligated to identify groups of goals, ascertain priorities, and reveal the dynamics of changing priorities. This is not to say that these things must happen before we do any training, nor is it to say that we must be as specific as a blueprint. But as part of the evaluation we must obtain some communicable representation of the different things different people want the training to accomplish.

Signs of Priority

The grand weakness in our present representation of goals is that we reveal few priorities, little ground for deciding which goals to pursue most vigorously. Our instructional technologists have ignored the problem, claiming responsibility only for already chosen goals. To read their literature is to learn that a goal unreached is a goal unsuitably pursued. That is no help. We will *continue* to

aspire for goals beyond our reach. A major responsibility of curriculum development is to assign priorities that indicate how much should be invested in the pursuit of each goal, and a major responsibility of curriculum evaluation is to point out less successful pursuits as a basis for reallocation of effort.

The notion of priorities is simple enough, but we have yet to represent them in operational language. To give real meaning to the term, we must choose among different implications for priority levels. My colleague, Tom Maguire, has acquainted me with several different implications of priority levels. I will mention two: priorities can imply either the initial assignment of resources or the rank order of guarantees of outcome. These two are surely correlated, but they do lead to different plans of action. The first definition says that greater effort will be allotted to higher priority goals. The second definition indicates that regardless of what initial emphasis is given to different goals, when formal or informal evaluation findings indicate that a high priority goal is not being satisfactorily pursued, other work will be dropped in favor of the high priority goal. One requires a careful plan, the other an effective monitoring device. Considerably more study needs to be given to operational definitions of the concept "priority among goals." We should be working toward the day when an outside evaluator could examine the goal specifications, priority lists, and progress reports and identify objectively the areas of under- and over-support.

A fourth advancement in language would be the development of more systematic rules for deriving teaching tactics from immediate goals and for deriving immediate goals from long-range goals. David Krathwohl (1965) has identified four levels of specificity of goals. Peter Taylor and Tom Maguire (1966) built a model to represent stages in the transformation of objectives, from societal press to terminal student behaviors. I would not claim that teaching tactics should be derived mechanically, but we should understand more about how appropriate tactics are selected and we should be in a position to compare tactics chosen intuitively with those obtained deductively.

Goals evolve over time; they should, of course. How can we distinguish between logical and capricious changes? In Figure 4 are my representations of the principal goals of Public Law 89-10 Title III as they have changed during the past few years. As an evaluator I should be able to say whether this restatement of purpose is a major or minor change, whether or not this representa-

President's Task Force on Education:	FL 89-10 Title III Purpose:	Title III Advisory Committee:	Nolan Estes, Assoc. Commissioner; Priority Funding (early 1967):	Revised Priorities (mid 1967):	Harold Howe, Commissioner; Primary Priority, FY 1968:
To stimulate and expand experimentation and innovation in education	To provide supplementary centers and services not now available	Fostering innovation, exemplary programs	Equalizing educational opportunities	Aid to deprived children in city core	Aid to deprived children in city core
	To create exemplary educational programs	Funding new uses of available facilities	Planning for metropolitan areas	Programs to advance "individualized instruction"	Secondary Priority, FY 1968:
		Multipurpose, high cost, visible projects	Meeting needs of rural communities	Exemplary programs of early childhood education	Better education in geographically isolated areas
			Coordinating all community resources	Quality education for minority groups	
				Better education in geographically isolated areas	
				Build planning and evaluation competence	

FIGURE 4. Evolution of Priorities for Title III Education Programs

tion is valid, and whether or not the change is logical. I cannot. At least one of my weaknesses is language.

There is a popular movie showing in the neighborhood theaters called *Cool Hand Luke*. Inside or outside prison, Luke is at cross purposes with the Establishment. He seals his fate by throwing back the warden's words:

"What we've got here is a failure to communicate."

That is just the way it is with educational evaluators. Course grades, test scores, federal project reports, accreditation standings, and research findings usually are hollow words and thoughtless patter to the educational decision maker. How can we tell him what he wants to know?

We have borrowed some of our better quantitative language from experimental psychology and from psychometric testing, a little from the school survey movement. We have many concepts from philosophy and the subject matter areas, many from the study of educational curricula. We need to develop some more for the express purpose of *telling others* about the antecedents, transactions, and outcomes of schooling. How can we build up this language? What should we do, for example, when some new languages like those in the National Assessment Program come along?

If we ask the rationalist for advice he will tell us to think through our needs for new language, to choose our indicators carefully, and to define our terms explicitly. He will say to hold work conferences, to talk it over among ourselves. He will tell us to invite the curriculum developer, teacher, and taxpayer to study our glossary, to learn our new language, perhaps even to help us evaluate it.

If we ask the empiricist for advice he will tell us to *try out* some new language. He will say to try lots of things, to see how they work with the people we really want to talk to. Try to figure out what the curriculum developer, teacher, and taxpayer mean when they use the same language. Use examples. Use illustrations. Let everyone share in the experience.

I wonder.

References

- David P. Ausubel. "An Evaluation of the BSCS Approach to High School Biology." *American Biology Teacher* 28: 176-86; 1966.
- J. Thomas Hastings. "Curriculum Evaluation: The Whys of the Outcomes." *Journal of Educational Measurement* 3: 27-32; 1966.

David R. Krathwohl. "Stating Objectives Appropriately for Program, for Curriculum, and for Instructional Materials Development." *Journal of Teacher Education* 12: 83-92; March 1965.

Thomas S. Kuhn. *The Structure of Scientific Revolution*. Chicago: University of Chicago Press, 1962.

Dan C. Lortie. "Rational Decision Making: Is It Possible Today?" *The EPIE Forum* 1:6-9; November 1967.

Jum Nunnally. *Tests and Measurements: Assessment and Prediction*. New York: McGraw-Hill Book Company, 1959.

Michael Scriven. "The Methodology of Evaluation." *American Educational Research Association Monograph Series on Curriculum Evaluation*. Volume 1. Chicago: Rand McNally and Company, 1967. pp. 39-83.

Robert E. Stake. "The Countenance of Educational Evaluation." *Teachers College Record* 68: 523-40; April 1967.

Robert E. Stake. "A Research Rationale for EPIE." *The EPIE Forum* 1: 7-15; September 1967.

Daniel L. Stufflebeam. "A Depth Study of the Evaluation Requirement." *Theory into Practice* 5: 121-33; 1966.

Peter A. Taylor and Thomas O. Maguire. "A Theoretical Evaluation Model." *The Manitoba Journal of Educational Research* 1: 12-17; June 1966.

Ralph W. Tyler. "Assessing the Progress of Education." *Phi Delta Kappan* 47: 13-16; September 1965.

Evaluation as Enlightenment for Decision Making

DANIEL L. STUFFLEBEAM

FOR the past 2½ years I have been deeply engaged in evaluation activities with personnel from local schools, state education departments, and the U.S. Office of Education. Those activities, for the most part, have involved efforts to evaluate projects funded under Title I and Title III of the Elementary and Secondary Education Act of 1965. This paper is based on those experiences and is an attempt to summarize some of my ideas about the kinds of evaluation which are needed in current programs of educational change.

The paper is divided into two parts. Part I is concerned mainly with determining the present state of the art in educational evaluation. In this part, I have attempted to describe current requirements for educational evaluation, to illustrate that educators have thus far been ineffectual in their attempts to meet these requirements, and to point out some possible reasons for poor evaluations in education. In Part II of the paper, I have attempted to conceptualize some alternative approaches to educational evaluation. This second part of the paper includes attempts to define evaluation in general terms, to sketch four evaluation strategies which I think have particular relevance to educational change activities, and to explicate the structure of evaluation design.

I want to emphasize that my formulations are largely untested and are therefore highly tentative. I sincerely hope that the reader will find these rough ideas worthy of examination. If any of them are found to be viable, I hope that others will help me, both during and after this working conference, to refine and extend them.

Part I: The State of the Art in Educational Evaluation

Education is becoming increasingly valued as a means to meet the social and economic as well as the intellectual needs of society. To fulfill this expanding role, educators are being asked to deal with many critical societal problems. These include inequality of opportunities among racial groups, de facto segregation, riots in our cities, disillusionment of youth, and school dropouts. Clearly, the rising trend of these problems must be curbed and pushed back for the welfare of our civilization. Education is thus being given a most urgent and difficult charge, and to meet this charge educators must mount many new and innovative efforts.

The Setting

To help educators meet their new responsibilities, society is annually providing billions of dollars through federal, state, and foundation programs to education agencies at all levels. Examples of increased support to education include the Elementary and Secondary Education Act of 1965, the Head Start Program, the Education Professions Act, and the Experienced Teacher Fellowship Program. Many industries are also developing education components, and soon we will probably see many new education-industry combines and consortia. Clearly, in addition to new responsibilities, education also has unprecedented opportunities to improve and expand its programs.

These opportunities, however, have also carried requirements that educators evaluate their new plans and programs. These requirements are especially evident in new federal assistance programs, e.g., Title I and Title III of the Elementary and Secondary Education Act. Here, the law explicitly states that fund recipients will make at least annual evaluation reports. As a consequence, many educators at all levels for the first time are having to cope with requirements for formal evaluation.

Such requirements for evaluation seem reasonable; and, in my judgment, they are long overdue. Funding agencies and the public have the right to know whether their huge expenditures for education are producing the desired effects. Even more important than this, educators themselves need evaluative information to provide rational bases for their decisions among alternative plans

and procedures. However, to justify requirements for evaluation is not to operationalize them. Educators must respond to the requirements, and they must do so effectively.

The Need for Better Educational Evaluations

Without question, educators are responding to requirements for evaluation. The multitude of evaluation reports now available from local schools, state education departments, regional educational laboratories, etc., demonstrates that educators are expending significant amounts of time, effort, and money to evaluate their programs. However, the increased activity alone has not met the need for effective evaluations. While educators have been busy doing evaluations, the fruits of their efforts have not provided the information needed to support decision making related to the programs being evaluated.

Many of the completed evaluation reports contain only impressionistic information. Though such information may be pertinent to the concerns of decision makers, it usually lacks the level of credibility required by decision makers to defend their decisions, and seldom can such information be of material use in making important decisions. A case in point is the first annual report for Title I of the Elementary and Secondary Education Act.¹ This report was highly important since it encompassed thousands of Title I projects throughout the nation. However, it fell far short of being a useful document, for it was almost devoid of hard data. On the other hand, it did contain many anecdotal accounts wherein persons who were responsible for conducting Title I activities stated that they felt that their program was being successful; and many of them speculated as to the reasons for the alleged successes. Though these anecdotes may have touched key issues related to the improvement of the billion dollar per year Title I program, decision makers in the Congress, the U.S. Office of Education, state education departments, and local school districts could hardly base important decisions on a few "possibly accurate" pieces of testimony.

The situation is not much different in Title III of the Elementary and Secondary Education Act. Title III staff members in the U.S. Office of Education have continuously ranked the quality of

¹ Public Law 89-10: The Elementary and Secondary Education Act of 1965, Title I.

Title III proposals on a five point scale for each of 15 criteria.² The criterion relating to evaluation has consistently been ranked near the "poor" end of the scale and lower than 13 of the other criteria—the exception being the criterion related to dissemination. Guba has also suggested that evaluation plans in Title III proposals are weak.³ Based on his analysis of 32 Title III proposals, Guba concluded:

It is very dubious whether the results of these evaluations will be of much use to anyone. They are likely to fit well, however, into the conventional school man's stereotype of what evaluation is: something required from on high that takes time and pain to produce but which has very little significance for action.⁴

Unlike the Title I and Title III evaluations already referred to, some evaluations provide for hard data. For example, the evaluation report for New York City's Higher Horizons Program⁵ used rigorous research procedures to compare the performance of an experimental group receiving the Higher Horizons Program with the performance of a control group which was matched to the experimental group on several counts. The basic conclusions contained in this nearly 300 page report were typical of findings for rigorous educational evaluations: "There were no significant differences." In sharp contrast, however, the report also noted that the teachers and principals who had been involved in the program said that it was making differences so significant that the program simply could not be abandoned.

Though the Title I, Title III, and Higher Horizons evaluations differed as to rigor, they were alike in one respect. None of them provided much help to the decision maker for improving the programs being evaluated. While I have cited only three examples of the deficiencies in current evaluations, I think they are sufficiently weighty ones to illustrate my point. In too many cases, evaluation

² These criteria are listed on pp. 70-71 of the current Title III guidelines: *A Manual for Project Applicants and Grantees*. Washington, D.C.: U.S. Office of Education, 1967.

³ Egon G. Guba, *Evaluation and the Process of Change*. Notes and Working Papers Concerning the Administration of Programs Authorized under Title III of Public Law 89-10, The Elementary and Secondary Education Act of 1965 as amended by Public Law 89-750, April 1967, p. 312.

⁴ *Ibid.*

⁵ J. Wayne Wrightstone et al. *Evaluation of the Higher Horizons Program for Underprivileged Children*. Cooperative Research Project No. 1124. New York: Bureau of Educational Research, Board of Education of the City of New York.

reports provide little or no help to decision makers, and thus decision making in and about education must remain an arduous endeavor.

Problems in Educational Evaluation

What is the explanation for this situation? Why is it that educators are failing to provide evaluations which are at the same time useful and scientifically respectable? Why is it that evaluations which adhere to classical research methods provide information which is of only limited help in making decisions about programs? And why do the typical "no significant difference" findings in so many of these evaluations contravene the experiences of those who are intimately involved in the programs?

One cannot answer these questions simply on the grounds that evaluation practice lags too far behind theory, or that there is a lack of effort on the part of educators to evaluate their programs. Further, it is not enough to note that evaluation testimony given by witnesses is not credible, or that typical findings of no significant differences are correct because "nothing in education ever makes a difference." Rather, I think the lack of adequate evaluation information persists because of several fundamental problems which must be solved before educators can improve their evaluations. These problems include a lack of trained evaluators, a lack of appropriate evaluation instruments and procedures, and a lack of adequate evaluation theory. In my judgment, the most basic of these problems is a lack of adequate theory or conceptualizations pertaining to the nature of evaluations which are needed to accommodate educational programs.

Clearly, the conceptual bases for evaluations are of fundamental importance. If these conceptions are faulty, then the evaluations which are based on them must also be faulty. Thus, it would seem highly important to identify and examine the efficacy of conceptualizations which underlie current needs for evaluation as well as educators' attempts to meet these needs. It will be useful to divide these conceptualizations into three classes and to consider each one separately. The three classes are:

1. Conceptions of the nature of the educational programs for which evaluations are needed, i.e., of the decision processes and associated information requirements which evaluations must serve

2. Conceptions of the nature of evaluation, in general, and as related to specific classes of educational programs
3. Conceptions of the structure of evaluation designs needed to conduct educational evaluations.

Problems in Defining Requirements for Educational Evaluations

First, let us examine problems involved in providing an adequate focus for educational evaluation studies. Obviously, to evaluate, one must know what is to be evaluated. Gaining knowledge of what is to be evaluated, however, is currently a difficult task at best. Current needs for educational evaluation have arisen due to programs and activities which are new to the field of education. Such activities involve responsibilities newly assigned to educators, new kinds of relationships among different kinds and levels of agencies, and a need for cooperative decision making about education among a variety of education and noneducation agencies. It should come as no shock if the evaluation theory which has traditionally been viewed as appropriate for education is found no longer to be adequate to meet the information requirements in new educational programs. Clearly, many of the new programs in education are dramatically different from those of the past, and our evaluations should probably be geared to answer questions which are much different from those questions they have answered in the past.

What we need, I think, are conceptualizations to account for decision processes and information requirements in new educational programs. Programs to improve education depend heavily upon a variety of decisions, and a variety of information is needed to make and support those decisions. Evaluators charged with providing this information must have adequate knowledge about the relevant decision processes and associated information requirements before they can design adequate evaluations. They need to have knowledge about the locus, focus, timing, and criticality of decisions to be served. At present no adequate knowledge of decision processes and associated information requirements relative to educational programs exists. Nor is there any ongoing program to provide this knowledge. In short, there are no adequate conceptualizations of decisions and associated information requirements or programs to produce them.

Problems in Defining Educational Evaluation

Next, let us attend to problems pertaining to the meaning of educational evaluation. Usually educators have defined evaluation as the science of determining the extent to which objectives have been achieved. The first step in operationalizing this definition is to state program objectives in behavioral terms. Then one must define and operationalize criteria for use in relating outcomes to the objectives. Operationalizing such criteria includes the specification of instruments for measuring outcomes and standards for use in assigning values to the measured outcomes.

Standards may be either in absolute or relative terms. An absolute standard might be that students on the average should achieve at least some specified score on a selected achievement test. A relative standard might be that the group of students receiving a new program should achieve scores on a selected achievement test which on the average are higher than scores achieved by an equivalent group of students which received some alternative program. Regardless of the type of evaluative standard used, the data from such studies are analyzed after a complete cycle of the program to determine the extent to which the objectives were achieved.

Evaluations based upon the above definition of evaluation yield data about gross total program effects and then only in retrospect. Such data are useful for making judgments about a project after it has run full cycle, but they certainly are not adequate to assist educators in the initial planning and in the actual carrying through of programs. At best, therefore, such evaluations provide an insufficient solution to the evaluation problems of educators who must plan and execute innovative programs.

The inadequacy of extant conceptions of evaluation is illustrated by the following excerpt from testimony pertaining to Title I evaluations, given before a Congressional committee by a citizens group in New York City:

We ask for amendments to render the required evaluations of Title I projects meaningful. The Act states that evaluations must be made, not that they be utilized in future planning. In New York City this year, projects were recycled before last year's evaluations were submitted. To be made more useful, evaluations should have built into them alternatives and the recommendations of the evaluator. What is now an expensive exercise should be made a function to provide service to local school boards having the responsibility for making policy based on experience.

American business would not survive if its consultants did not supply management with alternatives after reviewing the efficacy of programs.⁶

Here, the major concern seems to be that reports yielded by current evaluation programs are neither sufficiently specific nor timely to influence educational programs. Obviously, evaluations which do not meet at least these two criteria are of little use.

Problems in Designing Educational Evaluations

Finally, let us consider problems relating to the methodology of evaluation. If current conceptions of evaluation are not adequate for evaluating current educational activities, neither can extant designs be adequate. Existing means for evaluation have been developed to serve the ends of evaluation as these ends have been conceived traditionally.

The inadequacy of extant evaluation methodology is revealed when one examines the designs educators use to evaluate their programs. If they use a design at all, it typically is an experimental design. The fundamental concern of experimental design is that data which are produced be internally valid, i.e., unequivocal. Several conditions are necessary to meet this criterion. The units to be measured should be randomly assigned to treatment and control conditions. For example, a set of students might be partitioned randomly into two groups—one to receive a new program, the other to receive the school's present offering in the area to be served by the new program. Next, the treatment and control conditions must be applied and held constant throughout the period of the experiment, i.e., they must conform to the initial definitions of these conditions. The new or traditional program conditions could not be modified in process, since in that event one could not tell what was being evaluated.

Also, all students in the experiment must receive the same amount of the treatment to which they are assigned; and care must be taken so that students receiving one treatment are not contaminated by the other treatment. If contamination occurred, one could not tell what had caused what after the project was completed. Therefore, until an experiment is completed, one must resist the temptation to apply the successful activities of one con-

⁶ *Citizens' Committee for Children of New York, Inc., Newsletter*. Statement of Mrs. Nathan W. Levin, Chairman of the Educational Services Section before the Subcommittee on the Elementary and Secondary Education Act of the Education and Labor Committee of the House of Representatives, March 18, 1967.

dition to students receiving a different condition, even if the activities in the latter condition are obviously failing.

Finally, an instrument which is valid and reliable for the specified criterion variable must be administered after a certain period of time—usually a complete program cycle—to subjects from both parts of the experiment. Then, if all of the above conditions were met, one could use predetermined statistical procedures and decision rules to determine unequivocally that there were, or were not, significant differences between the experimental and control groups on the outcome variable of interest.

On the surface, the application of experimental design to evaluation problems seems reasonable, since traditionally both experimental research and evaluation have been used to test hypotheses about the effects of treatments. However, there are four distinct problems with this reasoning.

First, the application of experimental design to evaluation problems conflicts with the principle that evaluation should facilitate the continual improvement of a program. Experimental design prevents rather than promotes changes in the treatment because treatments cannot be altered in process if the data about differences between treatments are to be unequivocal. Thus, the treatment must accommodate the evaluation design rather than vice versa; and the experimental design type of evaluation prevents rather than promotes changes in the treatment.

It is probably unrealistic to expect directors of innovative projects to accept conditions necessary for applying experimental design. Obviously, they cannot constrain their treatment to its original definition just to ensure internally valid end-of-year evaluative data. Rather, project directors must use whatever evidence they can obtain to continually refine and sometimes radically change both the design and its implementation. It is thus contended here that conceptions of evaluation are needed which would result in evaluation programs which would stimulate rather than stifle dynamic development of programs.

A second flaw in the experimental design type of evaluation is that it is useful for making decisions after a project has run full cycle but almost useless as a device for making decisions during the planning and implementation of a project. It provides data after the fact about the relative effectiveness of two or more treatments. Such data, however, are neither sufficiently specific and comprehensive nor are they provided at appropriate times to assist

the decision maker in determining what a project should accomplish, how it should be designed, or whether the project activities should be modified in process. At best, experimental design evaluation reflects post hoc on whether a project did whatever it was supposed to do. At that time, however, it is too late to make decisions about plans and procedures which have already largely determined the success or failure of the project.

Guba⁷ has pointed out a third problem with the experimental design type of evaluation; it is well suited to the antiseptic conditions of the laboratory but not the septic conditions of the classroom. The potential confounding variables must be either controlled or eliminated through randomization if the study results are to have internal validity. However, in the typical educational setting this is nearly impossible to achieve. For example, consider the following quotation from an evaluation report completed by Julian Stanley:

Even if the program does have considerable cumulative influence on a person's career, this may be slow in appearing and so interactive with other influences that it cannot be discerned clearly by the person himself or by others.

Nevertheless, we must use whatever evidence that can be adduced to determine whether or not such programs are worth repeating and, if so, how they should be modified in order to be more effective. Ideally, in the experimental design sense, we should conduct the program as a controlled experiment, with a well-matched control group that does not attend the institute, and follow up both groups for quite a few years in order to determine how they diverge. If recruiting begins early enough and the applicant group is able enough to provide both groups at a sufficiently high level, this might be done, though the "reactivity" of the disheartened rejectees, the self-fulfilling prophecy of the rejectees, and the inability to control the summer activities of the rejectees might undesirably affect the outcome of the experiment. Merely having on one's record the fact of attending a certain prestigious program, like displaying one's Phi Beta Kappa key, might be a powerful aid. . . . Our chief way of evaluating the success of the program is via reports from staff and participants, particularly the latter.⁸

Egon G. Guba. "Methodological Strategies for Educational Change." Paper presented to the Conference on Strategies for Educational Change, Washington, D.C., November 8-10, 1965.

⁸ Julian Stanley. *Benefits of Research Design; A Pilot Study*. Final Report, Project No. X-005, Grant OE5-10-272, U.S. Department of Health, Education, and Welfare, U.S. Office of Education, Bureau of Research, August 1966.

γ

In the above quotation, Professor Stanley has pointed to many of the reasons why experimental design does not seem well suited to evaluation problems in education. In many innovative programs there clearly are a multitude of confounding factors which simply cannot effectively be controlled.

The existence of potentially confounding factors such as those named by Stanley gives rise to a fourth kind of problem inherent in the experimental design type of evaluation. *While internal validity may be gained through the control of extraneous variables, such an achievement is accomplished at the expense of external validity.* If the extraneous variables are tightly controlled, one can have much confidence in the findings pertaining to how an innovation operates in a controlled environment. However, such findings may not be at all generalizable to the real world where the so-called extraneous variables operate freely. Clearly, it is important to know how educational innovations operate under real world conditions.

Thus far, in this paper, I have attempted to depict the state of the art in educational evaluation. To begin with, I pointed out that educators are being faced with many new and different requirements for evaluation. Then I attempted to establish that educators' attempts to meet these requirements thus far have been ineffectual. Finally, I suggested that there are three types of conceptual problems which prevent educators from providing effective evaluations. These are:

1. A lack of understanding of decision processes and information requirements in current programs of educational change
2. The lack of a definition of educational evaluation which is pertinent to emergent requirements for educational evaluation
3. A lack of appropriate evaluation designs.

Part II: The Nature of Evaluation

Since this is a working document, I should probably stop with the definition of some of the current needs and problems. Readers could then examine my statement and modify or replace it. After we had achieved agreement as to what the real problems are, we could proceed to develop relevant solutions. However, I have been asked to expose some of my ideas regarding solutions for the current difficulties as I see them. Thus, in the remainder of this paper, I shall propose some alternative conceptions regarding the nature of educational evaluation.

This part of the paper is divided into four major sections. The first section is an attempt to define evaluation in general. Then, in Section 2, an attempt is made to analyze emergent programs of educational change and to identify the types of decisions for which evaluations are needed in these programs. Section 3 contains outlines of four strategies for evaluating educational programs, and the paper is concluded in Section 4 with an attempt to outline the structure of evaluation design.

The General Nature of Evaluation

A Rationale

If decision makers are to make maximum, legitimate use of their opportunities, they must make sound decisions regarding the alternatives available to them. To do this, they must know what alternatives are available and be capable of making sound judgments about the relative merits of the alternatives. This requires access to relevant information. Decision makers should, therefore, maintain access to effective means for providing this evaluative information. Otherwise, their decisions are likely to be functions of many undesirable elements. Under the best of circumstances, judgmental processes are subject to human bias, prejudice, and vested interests. Also, there is frequently a tendency to over-depend upon personal experiences, hearsay evidence, and authoritative opinion; and, surely, all too many decisions are due to ignorance that viable alternatives exist.

Clearly, the quality of programs depends upon the quality of decisions in and about the programs; the quality of decisions de-

depends upon decision makers' abilities to identify the alternatives which comprise decision situations and to make sound judgments of them; making sound judgments requires timely access to valid and reliable information pertaining to the alternatives; and the availability of such information requires systematic means to provide it. The processes necessary for providing this information for decision making collectively comprise the concept of evaluation. Given this rationale, I will now suggest a definition of evaluation.

Evaluation Defined

Generally, evaluation means the provision of information through formal means, such as criteria, measurement, and statistics, to provide rational bases for making judgments which are inherent in decision situations. To clarify this definition, it will be useful to define several key terms. A decision is a choice among alternatives. A decision situation is a set of alternatives. Judgment is the assignment of values to alternatives. A criterion is a rule by which values are assigned to alternatives, and optimally such a rule includes the specification of variables for measurement and standards for use in judging that which is measured. Statistics is the science of analyzing and interpreting sets of measurements. Measurement is the assignment of numerals to entities according to rules, and such rules usually include the specification of sample elements, measuring devices, and conditions for administering and scoring the measuring devices. Stated simply, evaluation is the science of providing information for decision making.

The methodology of evaluation includes four functions: *collection*, *organization*, *analysis*, and *reporting* of information. Criteria for assessing the adequacy of evaluations include *validity* (is the information what the decision maker needs?), *reliability* (is the information reproducible?), *timeliness* (is the information available when the decision maker needs it?), *pervasiveness* (does the information reach all decision makers who need it?), and *credibility* (is the information trusted by the decision maker and those he must serve?).

Evaluation in Fields Other Than Education

The concept of evaluation as defined above is general, since the assigning of values to alternatives is common to all forms of human thought and activity, and since men have always sought

to establish rational defensible bases for their judgments. However, there are many kinds of evaluation which meet the conditions of the above definition, but which nevertheless may be distinguished one from the other. For example, market research, cost benefit analysis, experimental design, objective testing, operational analysis, operations analysis, operations research, Program Evaluation and Review Technique, Program Planning and Budgeting System, quality control, and systems analysis all fit the general definition of evaluation given above.

Each of these modes of inquiry is the application of systematic means to aid in the assignment of values to the alternatives in decision situations. These different kinds of evaluation may be differentiated by the decision situations they serve, the settings within which the decisions are made, the kinds of tools and techniques used, the level of precision in the information collection and analytical modes, and the methodological skills of those who conduct the evaluations and those who are served by the evaluations. These substantive and methodological differences probab'y explain why different names have been given to these different forms of evaluation. For example, consider the following statement by Quade:

Evaluations undertaken to enable decision makers to choose among systems, to discover whether a given system would accomplish its objectives, or to set up a framework within which tests of a system could be prepared came naturally to be called "systems analysis."⁹

While Quade acknowledged that systems analysis is a form of evaluation, he also noted that the name "systems analysis" was derived from the nature of this form of evaluation.

Historical review of the more highly developed forms of evaluation listed above reveals that each was developed for relatively specific applications. Program Evaluation and Review Technique (PERT) was developed to aid the military in making decisions in the development of complex weapon systems. Systems analysis was developed to aid the military in making decisions in the development and implementation of military operations. Experimental design was especially useful for making judgments about the relative merits of agricultural products. And, initially, objective testing

⁹ Edward S. Quade, editor. *Analysis for Military Decisions*. Chicago: Rand McNally & Company, 1967. p. 4.

was utilized largely as an aid to the military in selecting men for military service.

Clearly, the development of each of these forms of evaluation was precipitated by critical decision-making needs; and these forms of evaluation were thus based upon the types of decisions to be served and the settings within which they were to be made. New approaches to evaluation were developed because extant approaches did not fit the decision-making requirements as precisely as needed, and because the decisions to be made could have serious consequences if wrong choices were made. Military decisions could affect the outcome of wars; thus, operations research, systems analysis, etc., were developed. Business decisions could result in profit, loss, or bankruptcy for thousands of stockholders; thus, cost benefit analysis and market research were developed.

Evaluation in Education

In the past, decisions about education have had effects less tangible than those in business, agriculture, and the military. Thus, there have not been pressures in education equivalent to those in other fields to motivate the development of highly specialized forms of evaluation to serve well-defined classes of educational decisions. Indeed, most educators would be hard pressed to identify and define the critical decision situations in education which merit specialized means for evaluation. It cannot be said, however, that education has been devoid of evaluation practices. Standardized testing has been developed to a high art to aid in college entrance decisions, the passing or failing of students, the assignment of diplomas and degrees, and the placement of students in educational programs. The *Buros Mental Measurements Yearbooks*¹⁰ have been developed to aid educators in the selection and use of tests. And, recently, Project EPIE (Educational Products Information Exchange)¹¹ has been developed to assist educators in selecting from among alternative products which are related to education. Generally, however, educators have failed to develop specialized means to aid their decisions about programs.

¹⁰ Oscar K. Buros. *The Buros Mental Measurements Yearbooks*. Highland Park, New Jersey: The Gryphon Press, 1949.

¹¹ *The EPIE Forum*. A Monthly Publication of the Educational Products Information Exchange Institute Created by and for Professionals in Education. New York: Educational Products Information Exchange Institute.

A prevalent position in education has been to avoid "reinventing the wheel," but instead to look to other fields in which problems similar to those in education have been faced and solved. This reasoning has led educators to adopt such evaluation modes as experimental design. Here a technique, previously utilized to assist farmers to select from among alternative kinds of fertilizer and seed, is being used to assist educators to select from among alternative educational innovations. The analogy between educational innovations and fertilizer is hopefully remote.

More recent forms of such borrowings are those of Program Evaluation and Review Technique, systems analysis, and the Program Planning and Budgeting System. At this point I would like to note that selective borrowing from other fields can save educators a great deal of time and effort. However, I also want to caution that wholesale, nonselective borrowing of techniques from other fields can result in the misapplication of techniques which never were intended for and do not fit educational situations. I think that educators' use of experimental design to evaluate innovative programs is an example of what can happen in the latter case. The use of experimental design in such applications has cost educators much time and effort without yielding much assistance for decision making.

As stated earlier in the paper, I think educators need some new basic conceptualizations to enable development of evaluation theory and methodology which have specific relevance to educational problems. In the previous section I have suggested a general rationale and definition for evaluation. Now I will attempt to derive a rationale and definition for evaluations in education.

A Rationale for Educational Evaluation

The Title I and Title III programs of the Elementary and Secondary Education Act of 1965 provide a comprehensive, timely context for deriving a rationale for educational evaluation. Virtually every school district in the nation is involved with one or both of these programs. The purposes of these programs respectively are to increase the educational attainment, experiences, and opportunities of disadvantaged children; and to increase the amount and quality of innovation in local education agencies. Both programs are national in scope, design, and broad control. They are coordinated and specifically controlled at the state level and are

implemented in local school districts. Together, they provide more than one billion dollars annually to local education agencies.

Figure 1 contains a conceptualization of the process and decision functions of evaluation as they may exist in federal assistance programs such as the Title I and Title III programs. A set of feedback control loops illustrates the relationships among local, state, and national evaluations of activities of federal assistance programs. In Figure 1, the loop at the right shows local school activities; the center loop, state activities; and the left loop, federal activities. Each loop contains a set of blocks, varied in shape, which represent the major evaluation functions.

Block 1 portrays the local school district's program. This is the local context from which needs for educational change emerge and within which the changes to meet these needs must ultimately occur. It includes the *inputs* of the system, i.e., the learners, curriculum, staff, organization, policies, finances, physical facilities, and school-community relations, and the *outputs* of the system, i.e., the cognitive, psychological, physical, and social functioning of its students and alumni.

To the right of Block 1, information collection is depicted by the first segment of curved line. This is a systematic collection at the local level of all information needed for later decisions at local, state, and federal levels.

Block 2 depicts the organization of information. Here, information would be coded according to predetermined categories, processed, keypunched, filed regularly, and retrieved as needed.

At Block 3, information organized at Block 2 would be analyzed according to decision-making requirements at local, state, and national levels and reported to local and state decision makers.

Block 4 denotes program decisions made at the local level. Local school decision makers to be served by the evaluation include the board of education, the school administration, project supervisors, teachers, and principals.

The decisions made at Block 4 would be implemented at Block 5, thus reactivating the cycle with frequent modification of the school program at Block 1. This cycle is continuous.

Returning to Block 3, evaluation reports for the state education department would be prepared annually by all public school districts in the state. At Block 6, the state education department would organize these reports into types of projects and combine information from similar projects. This information would then be analyzed

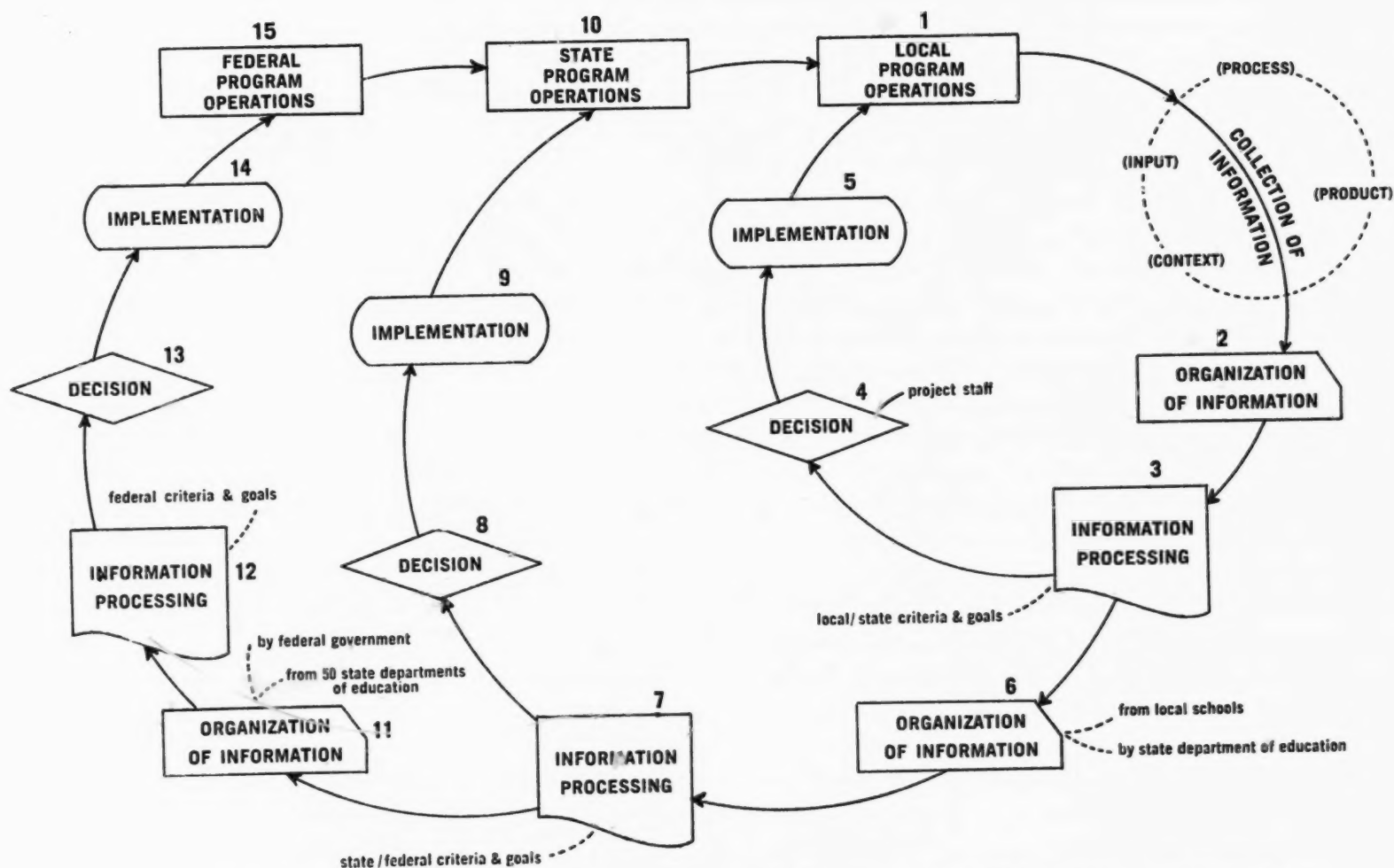


FIGURE 1. Feedback Control Loop: Evaluation in Federally Supported Educational Programs *

* Daniel L. Stufflebeam. "The Use and Abuse of Evaluation in Title III." *Theory Into Practice* 6 (3): 126-33; June 1967.

at Block 7 to determine the strengths and weaknesses of the statewide program. The state program officials would use this information to assess the statewide educational needs and problems to make decisions about program emphases and state control at Block 8. Decisions made at Block 8 would be implemented at Block 9, affecting the state program at Block 10, and reactivating the cycle at Block 1.

At Block 7, annual product evaluation reports from 50 states would be sent to the federal agency. This information would then be organized at Block 11, so that major program thrusts could be examined and analyzed on a nationwide basis at Block 12, and so that reports could be prepared for the Associate Commissioner for Elementary and Secondary Education, the Commissioner of Education, the Secretary of Health, Education, and Welfare, the President, and the Congress. Decisions about program emphases and funding would be made at the federal level at Block 13 and implementation of such decisions at Block 14 would affect the federal program at Block 15, the state program at Block 10, and the local school projects at Block 1, thus reactivating the cycle.

Summarized, Figure 1 demonstrates: (a) information for evaluation at federal, state, and local levels will be collected largely at the local level; (b) this information will form the basis for federal, state, and local decisions which will ultimately affect local operations; and (c) evaluation plans must be developed, communicated, and coordinated at federal, state, and local levels if the information schools provide is to be adequate for assisting in the decision process at each of these levels.

Obviously, to develop an appropriate evaluation system for programs such as Title I and Title III, one must first have some knowledge of the decision situations to be served. Optimally, such knowledge of decision situations should answer several questions. First, one should identify the *locus* of decision making, in terms of the level(s) at which authority and responsibility for decision making are vested, i.e., local, state, and/or national. Second, it is desirable to identify the *focus* of the decisions—are they related to goals of research, development, training, diffusion, etc.? Third, one needs knowledge of the *substance* of the decisions (are they related to mathematics, language arts, etc., and what are the alternatives in each decision situation?). Fourth, one needs to know the *function* of the decisions—are they for the planning, programming, implementing, or recycling of activities? Fifth, one needs knowl-

The Strategies

	Context Evaluation	Input Evaluation	Process Evaluation	Product Evaluation
OBJECTIVE	To define the operation context , to identify and assess needs in the context, and to identify and delineate problems underlying the needs.	To identify and assess system capabilities , available input strategies , and designs for implementing the strategies.	To identify or predict, in process, defects in the procedural design or its implementation, and to maintain a record of procedural events and activities .	To relate outcome information to objectives and to context, input, and process information.
METHOD	By describing individually and in relevant perspectives the major subsystems of the context; by comparing actual and intended inputs and outputs of the subsystems; and by analyzing possible causes of discrepancies between actualities and intentions.	By describing and analyzing available human and material resources, solution strategies, and procedural designs for relevance, feasibility and economy in the course of action to be taken.	By monitoring the activity's potential procedural barriers and remaining alert to unanticipated ones.	By defining operationally and measuring criteria associated with the objectives, by comparing these measurements with predetermined standards or comparative bases, and by interpreting the outcome in terms of recorded input and process information.
RELATION TO DECISION-MAKING IN THE CHANGE PROCESS	For deciding upon the setting to be served, the goals associated with meeting needs and the objectives associated with solving problems, i.e., for planning needed changes.	For selecting sources of support , solution strategies , and procedural designs , i.e., for programming change activities.	For implementing and refining the program design and procedure , i.e., for effecting process control.	For deciding to continue, terminate, modify or re-focus a change activity, and for linking the activity to other major phases of the change process, i.e., for evolving change activities.

FIGURE 2. The CIPP Evaluation Model—A Classification Scheme of Strategies for Evaluating Educational Change *

* Daniel L. Stufflebeam. "The Use and Abuse of Evaluation in Title III." *Theory Into Practice* 6 (3): 126-33; June 1967.

edge of the *objects* of the decisions (e.g., persons, places, events, or things?). Sixth, one obviously needs advance knowledge of the *timing* of decisions. And, finally, one needs knowledge of the relative *criticality* of decisions.

Considering all of the decision-making variables I have listed above, it is clear that one could identify many, many, different kinds of educational decision situations in education. Thus, it would also be possible to identify many different kinds of evaluation. However, it should prove more useful to develop a parsimonious classification system for kinds of educational evaluation which is intermediate between the general conceptual definition of evaluation given above and the many specific applied kinds of evaluation which could be derived from the use of all of the above-named variables in a detailed analysis and classification of education decision situations. Then it should be possible to derive useful names for the identified classes of educational evaluation.

To assist in developing a parsimonious classification system for educational decision situations in programs such as Title I and Title III, I have found it useful initially to focus exclusively on the functions of decisions.¹² I would postulate that functions of decision situations in education may be classified as *planning*, *programming*, *implementing*, and *recycling*. *Planning* decisions are those which focus needed improvements by specifying the domain, major goals, and specific objectives to be served. *Programming* decisions specify procedure, personnel, facilities, budget, and time requirements for implementing planned activities. *Implementing* decisions are those in directing programmed activities. *Recycling* decisions include terminating, continuing, evolving, or drastically modifying activities.

Four Strategies for Evaluating Educational Programs

Given these four kinds of educational decisions to be served, there are also four kinds of evaluation. These are portrayed in Figure 2 as context, input, process, and product evaluation. *Context* evaluation would be used when a project is first being planned. *Input* evaluation would be used immediately after context for specific programming of activities. *Process* evaluation would be used

¹² Daniel L. Stufflebeam. "The Use and Abuse of Evaluation in Title III." *Theory Into Practice* 6 (3): 126-33; June 1967.

continuously during the implementation of the project. *Product* evaluation would most likely be used after a complete cycle of the project. Each of these kinds of evaluation will be considered individually.

Context Evaluation

The major objective of context evaluation is to define the environment where change is to occur, the environment's unmet needs, and the problems underlying those needs. For example, the environment may be defined as the inner city elementary schools of a large metropolitan area. Study of such a setting might reveal that the actual reading achievement levels of children in this area are far below what the school system expects for them. This would be the identification of a need, i.e., the context evaluation would have revealed that the children's reading achievement levels need to be raised.

As a next step in the context evaluation, the school would attempt to identify the reasons for such a need. Are the students receiving adequate instruction? Are the instructional materials appropriate for them? Is there a major language barrier? Is there a high incidence of absenteeism? Is the school's expectation for these students reasonable? These are what I mean by potential problems. They are potential dilemmas which prevent the achievement of desired goals and thereby result in the existence of needs.

The method of context evaluation begins with a conceptual analysis to identify and define the limits of the domain to be served as well as its major sub-parts. Next, empirical analyses are performed, using techniques such as sample survey, demography, and standardized testing. The purpose of this part of context evaluation is to identify the discrepancies among intended and actual situations for each of the sub-parts of the domain of interest, and thereby to identify needs. Finally, context evaluation involves both empirical and conceptual analyses, as well as appeal to theory and authoritative opinion, to aid judgments regarding the basic problems underlying each need.

Decisions served by context evaluation include deciding upon the setting to be served, the goals associated with meeting needs, and the objectives associated with solving problems. Such decisions usually appear in the introductory sections of proposals to funding agencies or in requests for proposals by funding agencies.

Input Evaluation

To determine how to utilize resources to meet program goals and objectives, it is necessary to do an input evaluation. Its objective is to identify and assess relevant capabilities of the proposing agency, strategies which may be appropriate for meeting program goals, and designs which may be appropriate for achieving objectives associated with each program goal. The end product of input evaluation is an analysis of alternative procedural designs in terms of potential costs and benefits.

Specifically, alternative designs are assessed in terms of their resource, time, and budget requirements; their potential procedural barriers; the consequences of not overcoming these barriers; the possibilities and costs of overcoming them; relevance of the designs to program objectives; and overall potential of the design to meet program goals. Essentially, input evaluation provides information for deciding whether outside assistance should be sought for meeting goals and objectives; what strategy should be employed, e.g., the adoption of available solutions or the development of new ones; and what design or procedural plan should be employed for implementing the selected strategy.

Methods for input evaluation are lacking in education. The prevalent practices include committee deliberations, appeal to the professional literature, and the employment of consultants. In a few areas, formal instruments exist to aid decision makers in making input decisions. In the design of testing programs, one may obtain substantial help by referring to the *Buros Mental Measurements Yearbooks*.¹³

The educational researcher, who wants to select an experimental design, can receive material assistance in identifying and assessing alternative experimental designs by referring to the Campbell-Stanley chapter on experimental design in Gage's *Handbook of Research on Teaching*.¹⁴ In this chapter, the decision situation posed to the researcher in need of an experimental design is neatly laid out in the form of alternative designs which are relevant to experimental research. Each of these designs is rated regarding its potential to meet criteria of internal and external validity. Further,

¹³ Buros, *op. cit.*

¹⁴ N. L. Gage, editor. *Handbook of Research on Teaching*. The American Educational Research Association. Chicago: Rand McNally & Company, 1963.

procedural barriers or sources of invalidity are identified for each of the listed designs.

Decisions based upon input evaluation usually result in the specification of procedures, materials, facilities, schedule, staff requirements, and budgets in proposals to funding agencies. From the information provided in the proposals, the funding agencies in turn do an input evaluation to determine whether or not to fund the proposed projects. Funding agencies commonly employ expert consultants to serve as judges in their input evaluations.

Process Evaluation

Once a designed course of action has been approved and implementation of the design has begun, process evaluation is needed to provide periodic feedback to project managers and others responsible for continuous control and refinement of plans and procedures. The objective of process evaluation is to detect or predict, during the implementation stages, defects in the procedural design or its implementation. The overall strategy is to identify and monitor, on a continuous basis, the potential sources of failure in a project. These include interpersonal relationships among staff and students; communication channels; logistics; understandings of and agreement with the intent of the program by persons involved in and affected by it; adequacy of the resources, physical facilities, staff, and time schedule; etc.

As opposed to experimental design evaluation, process evaluation does not require control over assignment of subjects to treatments, nor that the treatments be held constant. Its purpose is to assist project personnel to make their decisions a bit more rational in their continual efforts to improve the quality of the program. Thus, under process evaluation, the evaluator accepts the program as it is and as it evolves, and monitors the total situation as best he can by focusing the most sensitive and nonintervening data collection devices and techniques that he can obtain on the most crucial aspects of the project. Such evaluation is multivariate, and not all of the important variables can be specified before a project is initiated. The process evaluator focuses his attention on theoretically important variates, but he also remains alert to any unanticipated but significant events. Under process evaluation, information is collected daily, organized systematically, analyzed periodically, e.g., weekly, and reported as often as project personnel require such information, e.g., monthly.

Thus, project decision makers are provided not only with information needed for anticipating and overcoming procedural difficulties, but also with a record of process information to be used later for interpreting project outcomes.

Product Evaluation

Product evaluation is used to determine the effectiveness of the project after it has run full cycle. Its objective is to relate outcomes to objectives and to context, input, and process, i.e., to measure and interpret outcomes.

The method is to operationally define and measure criteria associated with the objectives of the activity, to compare these measurements with predetermined absolute or relative standards, and to make rational interpretations of the outcomes using the recorded context, input, and process information. Criteria for product evaluation may be either instrumental or consequential, a distinction pointed out earlier by Scriven.¹⁵ Instrumental criteria are related to program outcomes which contribute to the achievement of behavioral objectives. Clark and Guba¹⁶ have developed a taxonomy of instrumental objectives and associated criteria which are related to educational change. My adaptation of their scheme is presented as Figure 3. Consequential criteria are primarily those pertaining to behavioral objectives. Bloom's *Taxonomy of Educational Objectives*¹⁷ is useful in the identification of consequential objectives.

In the change process, product evaluation provides information for deciding to continue, terminate, modify, or refocus a change activity, and for linking the activity to other phases of the change process. For example, a product evaluation of a program to develop after-school study for students from disadvantaged homes might show that the development objectives have been satisfactorily achieved and that the developed innovation is ready to be diffused to other schools which need such an innovation.

¹⁵ Michael Scriven. *The Methodology of Evaluation*. Bloomington, Indiana: Indiana University, Social Science Education Consortium, Publication #110, 1965.

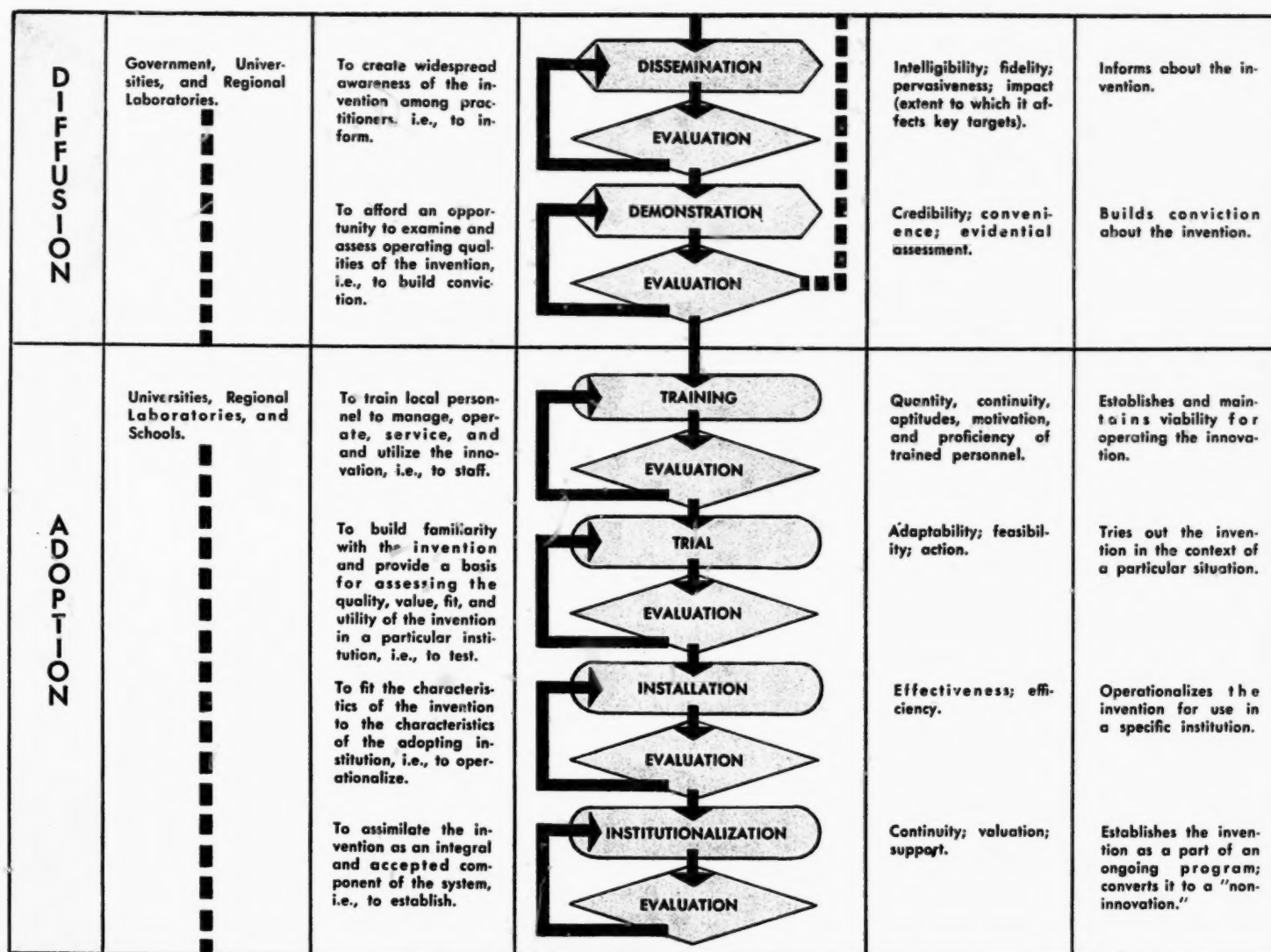
¹⁶ David L. Clark and Egon G. Guba. "An Examination of Potential Change Roles in Education," Paper read at the Symposium on Innovation in Planning School Curricula, Airlie House, Virginia, October 1965.

¹⁷ Benjamin S. Bloom. *Taxonomy of Educational Objectives: The Classification of Educational Goals, Handbook I: Cognitive Domain*. New York: Longmans, Green and Company, Inc., 1956.

FIGURE 3. A Process Chart Depicting the Role of Evaluation in the Change Process *

* Based upon "A Classification Scheme of Processes Related to and Necessary for Change in Education" by David L. Clark and Egon G. Guba. Chart reprinted from "A Depth Study of the Evaluation Requirement," Daniel L. Stufflebeam. *Theory Into Practice* 5: 121-33; June 1966.

	AGENCY	OBJECTIVE	PROCESS	CRITERIA	RELATION TO CHANGE
RESEARCH	Universities, Research and Development Institutions, and Regional Laboratories.	To advance knowledge, i.e., to depict, correlate, conceptualize, and test.		Validity (internal and external).	Provides basis for invention.
DEVELOPMENT	Universities, Research and Development Institutions, Regional Laboratories, and Industries.	To formulate a new solution to an operating problem or to a class of operating problems i.e., to innovate. To draft a plan for constructing the innovation, i.e., to construct the blue-print. To build the components, i.e., to construct. To integrate the components into an operating system, i.e., to finalize for marketing.		Face validity (appropriateness); estimated viability; impact (relative contribution). Feasibility (production and utilization); tractability (ease of managing, controlling, and instructing in the use of). Design specifications; individual performance. Design specifications; total performance; viability; efficiency.	Produces the invention. Engineers the invention to fit the characteristics of the target situation. Produces the components necessary for implementing the design. Produces the coordinated operating system.



Given these four kinds of evaluation, it is next necessary to consider methodology for implementing them. This problem is considered in the next section of this paper.

The Structure of Evaluation Design

Once an evaluator has selected an evaluation strategy, e.g., context, input, process, or product, he must next select or develop a design to implement his evaluation. This is a difficult task, since few generalized evaluation designs exist which are adequate to meet emergent needs for evaluation. Thus, educators must typically develop evaluation designs *de novo*.

The remainder of this paper is an attempt to provide a general guide for developing evaluation designs. Specifically, I will attempt to define design in general terms and to explicate the general structure of designs for educational evaluation. Hopefully, this general treatment of evaluation design will be of some help to educators in ordering their minds as they approach problems of designing evaluations. Also, I am hopeful that the following material might stimulate methodologists who are more capable than I to develop generalized designs for context, input, process, and product evaluation.

Design Defined

In general, design is the preparation of a set of decision situations for implementation toward the achievement of specified objectives. This definition says three things. First, one must identify the objectives to be achieved through implementation of the design. In a product evaluation, for example, such an objective might be to make a determination of whether all students in a remedial reading program attained specified levels of specific reading skills. Second, this definition says that one should identify and define the decision situations in the procedure for achieving the evaluation objective. For example, in the remedial reading case cited above, one would want to identify the available measuring devices which might be appropriate for assessing the specified reading skills. Third, for each identified decision situation the evaluator needs to make a choice among the available alternatives. Thus, the completed evaluation design would contain a set of decisions as to how the evaluation is to be conducted and what instruments will be used.

It should be useful to evaluators to have available a list of the decision situations which are common to many evaluation designs.

This would enable them to approach problems of evaluation design in a systematic manner. Further, such a list could serve as an outline for the content of evaluation sections in research and development proposals. Funding agencies should also find such a list useful in structuring their general guidelines for evaluations which they provide to potential proposal writers. Also, such a list should be useful to training agencies for defining the role of the evaluation specialist.

Figure 4 is an attempt to provide such a general list of decision situations for evaluation designs. By presenting this general list, I am asserting that the structure of evaluation design is the same for context, input, process, or product evaluation. This structure includes six major parts. These are (a) focusing the evaluation, (b) information collection, (c) information organization, (d) information analysis, (e) information reporting, and (f) the administration of evaluation. Each of these parts will be considered separately.

Focusing the Evaluation

The first part of the structure of evaluation design is that of focusing the evaluation. The purpose of this part is to spell out the ends for the evaluation and to define policies within which the evaluation must be conducted. Specifically, this part of evaluation design includes four steps.

The first step is to identify the major levels of decision making for which evaluation information must be provided. For example, in the Title III program of the Elementary and Secondary Education Act, evaluative information from local schools is needed at local, state, and national levels. It is important to take all relevant levels into account in the design of evaluations, since different levels may have different information requirements and since the different agencies may need information at different times.

Having identified the major levels of decision making to be served by evaluation, the second step is to identify and define the decision situations to be served at each level. Given our present low state of knowledge about decision making in education, this is a very difficult task. However, it is also a very important one and should be done as well as is practicable. First, decision situations should be identified in terms of those responsible for making the decisions, e.g., teachers, principals, board of education members, and state legislators. Next, major types of decision situations should

The logical structure of evaluation design is the same for all types of evaluation, whether context, input, process, or product evaluation. The parts, briefly, are as follows:

A. *Focusing the Evaluation*

1. Identify the major level(s) of decision making to be served, i.e., local, state, and/or national.
2. For each level of decision making, project the decision situations to be served and describe each one in terms of its locus, focus, criticality, timing, and composition of alternatives.
3. Define criteria for each decision situation by specifying variables for measurement and standards for use in the judgment of alternatives.
4. Define policies within which the evaluation must operate.

B. *Collection of Information*

1. Specify the source of the information to be collected.
2. Specify the instruments and methods for collecting the needed information.
3. Specify the sampling procedure to be employed.
4. Specify the conditions and schedule for information collection.

C. *Organization of Information*

1. Provide a format for the information which is to be collected.
2. Designate a means for coding, organizing, storing, and retrieving information.

D. *Analysis of Information*

1. Select the analytical procedures to be employed.
2. Designate a means for performing the analysis.

E. *Reporting of Information*

1. Define the audiences for the evaluation reports.
2. Specify means for providing information to the audiences.
3. Specify the format for evaluation reports and/or reporting sessions.
4. Schedule the reporting of information.

F. *Administration of the Evaluation*

1. Summarize the evaluation schedule.
 2. Define staff and resource requirements and plans for meeting these requirements.
 3. Specify means for meeting policy requirements for conduct of the evaluation.
 4. Evaluate the potential of the evaluation design for providing information which is valid, reliable, credible, timely, and pervasive.
 5. Specify and schedule means for periodic updating of the evaluation design.
 6. Provide a budget for the total evaluation program.
-

FIGURE 4. Developing Evaluation Designs

be identified, e.g., appropriational, allocational, approval, or continuation. Then these types of decision situations should be classified by focus, e.g., research, development, diffusion, or adoption. (This step is especially helpful toward identifying relevant evaluative criteria.)

These identified decision situations should then be analyzed in terms of their relative criticality. In this way, relatively less important decisions which would expend evaluation resources needlessly can be eliminated from further consideration. Next, the timing of the decision situation to be served should be estimated so that the evaluation can be geared to provide relevant data prior to the time when decisions must be made. And, finally, an attempt should be made to explicate each important decision situation in terms of the alternatives which may reasonably be considered in reaching the decision.

Once the decision situations to be served have been explicated, the third step is to define relevant information requirements. Specifically, one should define criteria for each decision situation by specifying variables for measurement and standards for use in the judgment of alternatives.

The fourth and final step in focusing the evaluation is to define policies within which the evaluation must operate. For example, one should determine whether a "self evaluation" or "outside evaluation" is needed. Also, it is necessary to determine who will receive evaluation reports and who will have access to them. Finally, it is necessary to define the limits of access to data for the evaluation team.

Collection of Information

The second major part of the structure of evaluation design is that of planning the collection of information. This section obviously must be keyed very closely to the criteria which were identified in the Evaluation Focus part of the design.

Using those criteria, one should first identify the sources of the information to be collected. These information sources should be defined in two respects: first, the origins for the information, e.g., students, teachers, principals, or parents, and second, the present state of the information, e.g., in recorded or nonrecorded form.

Next, one should specify instruments and methods for collecting the needed information. Examples include achievement tests,

interview schedules, and searches through the professional literature. Metfessel and Michael¹⁸ have recently provided a comprehensive list of instruments with potential relevance for data collection in evaluations.

For each instrument that is to be administered, one should next specify the sampling procedure to be employed. Where possible, one should avoid administering too many instruments to the same person. Thus, sampling without replacement across instruments can be a useful technique. Also, where total test scores are not needed for each student, one might profitably use multiple matrix sampling where no student attempts more than a sample of the items in a test.

Finally, one should develop a master schedule for the collection of information. This schedule should detail the interrelations between samples, instruments, and dates for the collection of information.

Organization of Information

A frequent disclaimer in evaluation reports is that resources were inadequate to allow for processing all of the pertinent data. If this problem is not to arise, one should make definite plans regarding the third part of evaluation design: organization of information. Organizing the information that is to be collected includes providing a format for classifying information and designating means for coding, organizing, storing, and retrieving the information.

Analysis of Information

The fourth major part of evaluation design is analysis of information. The purpose of this part is to provide for the descriptive or statistical analyses of the information which is to be reported to decision makers. This part also includes interpretations and recommendations. As with the organization of information, it is important that the evaluation design specify means for performing the analyses. The role should be assigned specifically to a qualified member of the evaluation team or to a special agency which specializes in doing data analyses. Also, it is important that

¹⁸ Newton S. Metfessel and William B. Michael. "A Paradigm Involving Multiple Criterion Measures for the Evaluation of the Effectiveness of School Programs." *Educational and Psychological Measurement* 27: 931-36; 1967.

those who will be responsible for the analysis of information participate in designing the analysis procedures.

Reporting of Information

The fifth part of evaluation design is the reporting of information. The purpose of this part of a design is to ensure that decision makers will have timely access to the information they need and that they will receive it in a manner and form which facilitate their use of the information. In accordance with the policy for the evaluation, audiences for evaluation reports should be identified and defined. Then means should be defined for providing information to each audience. Subsequently, the format for evaluation reports and reporting sessions should be specified. And, finally, a master schedule of evaluation reporting should be provided. This schedule should define the interrelations between audiences, reports, and dates for reporting information.

Administration of Evaluation

The last part of evaluation design is that of administration of the evaluation. The purpose of this part is to provide an overall plan for executing the evaluation design. The first step is to define the overall evaluation schedule. For this purpose, one might usefully employ a scheduling technique such as Program Evaluation and Review Technique. The second step is to define staff requirements and plans for meeting these requirements. The third step is to specify means for meeting policy requirements for conduct of the evaluation. The fourth step is to evaluate the potential of the evaluation design for providing information which is valid, reliable, credible, timely, and pervasive. The fifth step is to specify and schedule means for periodic updating of the evaluation design. And the sixth and final step is to provide a budget for the evaluation.

Finally, I have reached the end of my paper. While I have only scratched the surface regarding educational evaluations, it is clear to me that the design and analysis of educational evaluation is a most complex and difficult undertaking. Surely, all of us who are committed to reshaping the world of educational evaluation must work very, very hard if we are to make any progress. If progress is not made in this area, I am convinced that education will be a casualty for want of adequate information to support vital decisions in and about education.

Emotion: The Missing Link in Education¹

WALCOTT H. BEATTY

MY CONCERN for the past 20 years as a teacher and as a research psychologist has been to understand how learning takes place in real life situations. Theories of learning which have been developed in the laboratory with animals and with the study of oversimplified human learnings are plainly not adequate to the task of answering this question. To date, our best clues have come from clinical psychologists, and educators have tended to respond favorably to the writings of Carl Rogers, Abraham Maslow, and Arthur Combs. However, as I have viewed it, this favorable response has led to little change in what is actually happening in the classroom.

This phenomenon has puzzled me and I have given much thought to the problem. As I have studied it and experimented in my own teaching I have gradually come to a conclusion. As educators we have devoted almost exclusive attention to intellectual and cognitive processes and their effective development. Men like Rogers and Combs are talking about the role of feelings and emotion in behavior and how their development can be fostered effectively. We have heard them say that, for learning to take place, a teacher must be accepting of children, must be understanding, and must be open and transparent in relationships with children, but somehow we seem to have missed the point. *We have tried to apply these guidelines to the fostering of intellectual behavior in children, when*

¹ Adapted from a speech given at the Conference on Issues in Human Development, Present and Future, at the Institute for Child Study, University of Maryland, April 20, 1968.

the real point is that these guidelines must be applied in relation to the affective and emotional behavior of children.

We have been blind to this for the simple reason, I believe, that we distrust many of our own feelings and emotions. We do not understand the relationship between feelings and intellectual behavior. We are afraid or at least dismayed or embarrassed by the appearance of strong emotions in others or in ourselves.

There has been little written in the professional literature of education that explores the problem of educating the emotions. In 1938 Daniel Prescott published a book entitled *Emotion and the Educative Process* (1938). The book has, in my opinion, become a classic. It demonstrates clearly that research supports the idea that feelings and emotion play a critical role in blocking and in enhancing learning. Further, they are a major determinant of what will be learned in any situation. Prescott explores brilliantly the implications of these findings for education. Research since that time has continued to support Prescott's findings, and yet the area of feelings and emotion is almost totally neglected in our current educational processes.

A major development since the time that Prescott wrote his book, and clearly he was instrumental in this change, is that it is no longer possible to talk about feelings and emotion separately from a consideration of the total functioning organism. Two of the most recent books summarizing what is known about emotion are, of necessity, combination titles. Paul Young, who did his first writing in this field as long ago as did Prescott, has entitled his latest book *Motivation and Emotion* (1961). Magda Arnold, who has produced another comprehensive and deeply thoughtful book about emotion, has entitled her two-volume work *Emotion and Personality* (1960). I think it is clear that there are limitations to the study of any human function in isolation from total organismic functioning in environment.

In recent years Rodney Clark and I have published a "Self Concept Theory of Learning" (1962), which attempts to integrate personality theory, learning, and motivation into a single theory to explain the development of complex behavior and of behavior change. There are a number of inadequacies in the theory, but it illustrates that such an approach is productive. In this paper I would like to extend the theory by relating it to feelings and emotion. I will begin by giving some essential elements of the self concept theory and then discuss the role of feelings and emotion

in this explanation of behavior. Finally, I would like to speculate on how education might promote affective development in the light of these ideas.

Self Concept, Motivation, and Learning

The self concept is an organization of images which each person has about himself in the world. These images develop over time from the reflected appraisals of others around him. As he is reacted to as a boy, with all of the rich ramifications expected of a boy, he builds this into his self-image. As he is loved, he comes to see himself as lovable. These images also include the world around him, so that the image is one of a boy among people who are both male and female and as one who is loved by loving people. These images multiply and develop many refinements and complexities over time. This perceived-self-in-the-world is the only reality he can know, and his behavior is of necessity consistent with these perceptions. He will come to know about other ways of behaving, but these are his *own* ways of behaving because this is the kind of person he is.

One of the important things about being reared in a family is that for the most part it provides a relatively consistent set of appraisals which are reflected to the child. Therefore, the images he is building are reinforced and become, along with his behavior, satisfying. As the child grows older he experiences divergent appraisals of what he is like, but since they are inconsistent with the reality he has learned, these appraisals are initially resisted and screened out of awareness. If the new and divergent appraisals continue to bombard him and if there is insufficient reinforcement for his former images of self, then he will slowly change and integrate the new appraisals into self. Thus, as the individual grows and comes into contact with larger segments of the world, there is gradual and continuing change.

This description certainly captures some of the process of developing a self concept and of the slow change which takes place over time, but there is something missing. It makes the individual sound too passive. The active, interacting human being is much more dynamic. There is another part of the self concept which grows right along with the perceived self. Not only does a child experience appraisals of what he is like, but also he is appraised in terms of what he could or should be like. More important than this,

he has the models of mother and father who appear to be so much more effective than he is.

This experience leads to the development of a part of the self concept which I have called the concept of adequacy, the way the child perceives that he should be if he is really going to be adequate and effective in the world. These two parts, the perceived self and the concept of adequacy, make up the self concept. To some extent they overlap, so that the way the individual sees himself and the way he should be, to be adequate, are the same. However, to a large degree, they are different. It is this discrepancy between the two which is the source of motivation. An individual is continually striving to become more like his picture of an adequate self.

As the individual is aware of discrepancies between the perceived self and his picture of adequacy, he formulates goals, that is, things which he could do in the world which would decrease the discrepancies. He tries to realize these goals in whatever situation he encounters by taking actions in the direction of these goals. These actions are responded to by others in the situation and the individual must evaluate the meanings of these responses. If the responses are relevant to his perceived self and approximately consistent with the way in which he views his self, these responses are then evaluated to determine whether or not they indicate that he is becoming more adequate. Consistent responses which tell him that he is becoming more adequate stabilize the new behavior and lead to a change in the perceived self. He becomes more like his concept of adequacy, and this change is what we call learning.

To make it a little more concrete, let us say that a child sees himself as having few friends in a classroom. His concept of adequacy is that of one who has many friends. In the classroom he plays with a child he has not known well before, and the child asks him to come to his house after school. This reaction from the other child is consistent with his picture of himself as likeable and seems to indicate that he now has one more friend. His behavior of playing with a new child has been effective, and he will try it again. As long as it appears to work he will use it and gradually his perceived self will change to that of a person who has many friends.

Organizing Centers for Self and Adequacy

The concept of self theory in this form helps to clarify the source of motivation in behavior and makes clear that learning

must involve a change in self if it is to persist. However, this may not be too helpful to a teacher working with a particular child, until the teacher gets to know him very well. From my experience, reading, and research, I think it is possible to make a further breakdown in the nature of the self concept concerning the way in which it is organized. It appears to me that there are four organizing centers or nodal points around which perceptions of self and adequacy are clustered. I am postulating that every human being, regardless of his culture, has organized his experiences and learnings around four areas: worth, coping, expressing, and autonomy. Let me discuss each one briefly.

Worth. The experience of love, of being included, of being given priority over other things builds a child's picture of his worth. He also experiences others, his mother and father, for example, who are also loved and are included and are given priority. These others necessarily appear to have greater worth than his own as his parents do things which do not include him, and their needs are sometimes given priority, so that the child develops a picture of even greater worth which he may achieve in time.

This provides the motivation to become more like his parents in order to become more worthy. He seeks the ways in which he can get more love, be included more, and get more priorities. If the discrepancy between perceived self and adequacy is small, he is weakly motivated to change and appears secure. If the discrepancy is great, he appears jealous of others and strives ever harder for attention.

Coping. The experience of inability to do something, and then learning how, builds a child's picture of himself as a coping person. Again, his models—mother, father, older siblings—are so much more capable that he builds a picture of coping ability he would like to achieve. The amount and kinds of coping ability necessary in our society are so great that we have schools to provide help in the necessary learning. A child in a family where things are done together and where the child is helped to learn things goes to school with interest and motivation.

Before going on to the next area, let me comment that, although I have presented worth and coping as two separate organizing centers in self, they can become greatly entangled. Parents who offer love as the reward for "good" behavior and withhold it when the child fails to please them provide experiences which entangle and confuse these two areas. This develops the concept that

worth depends upon doing what others wish. Which of us is free of the irrational desire to please and be liked by almost everybody? The schools, by their use of grades as a kind of global evaluation of the child, contribute greatly to this fusing of worth and coping. It may be productive for individuals to strive continually for advancement in order to prove their worth, but "progress" comes at a high psychic cost.

Expressing. The nature of the organism is such that most sensations are experienced with an affective tone. They are pleasant or unpleasant. Those which are pleasant are sought after. This lays the basis for participating, either actively or passively, in the arts. Music, rhythm, painting, color, and many other experiences evoke pleasant feelings even though their enjoyment contributes nothing to being able to cope better. Our frenetic culture, with its stress on increasing coping ability and the perverted achievement of worth through coping, has almost destroyed the arts, and most people have only a large blind spot in this area. The schools provide experience in music, painting, and body movement in the first few grades, but these activities disappear by the fourth grade and return again only as electives in high school and college.

There is another aspect to this area which can lead to far more serious consequences than a mere blind spot. This is the fact that many of the things we are aware of around us stir up feelings and emotion, both pleasant and unpleasant, which we are forbidden to express. Our culture generally discourages the expression of strong emotions. Parents, teachers, and friends all seem to find them disturbing and would rather not be with us when we are angry or grief-stricken or even when we are exuberantly happy. This makes expression difficult and prevents our learning effective ways of expressing emotion.

Autonomy. As an individual grows and develops feelings of worth, ability to cope, and ability to express, he finds that every situation provides a stage with more alternatives open to him. As he discovers the alternatives which give him greater feelings of satisfaction, he becomes more autonomous, more capable of making choices and controlling his own future. Experiencing situations in which he is given independence and responsibility promotes such development.

It is the development of these four areas which gives meaning to the idea of maturity. My definition of a truly mature person is

one who feels worthy without having to defend his actions, feels confident that he will be able to cope with the situations which he is likely to face, can express himself so that he feels satisfaction and stays relatively free of tension and anxiety, and feels that every situation provides genuine choices through which he can affect his own future. The mature individual has not resolved all internal discrepancies between his perceived and adequate self, but his progress has greatly decreased the motivation for self enhancement, and he tends to turn toward working on the discrepancies in society which interfere with the development and functioning of mature individuals.

This has been, at most, a sketchy introduction to a self-concept theory of learning, but perhaps it lays the ground for further comments on feelings and emotion. The essence of the theory is that an individual interprets his experience according to his picture of himself in the world. His reactions to this experience are guided by his motivation to become a more adequate person. The consequences which his reactions bring are again evaluated and a new reaction appears, and so the cycle continues throughout life.

Feelings and Emotion

Now, how do feelings and emotion come into this process? It has already been necessary to use the terms in the discussion of the area of expressing and in the comments on maturity. Perhaps we can move ahead by stating two hypotheses. First, I propose that feelings and emotion can only be understood as they are related to a personality theory such as self concept. Second, I propose that feelings and emotion are two separate things. There are characteristics which define emotion which are not present in feelings when the term is used appropriately.

In Prescott's writing of 1938 he discussed feelings and emotion separately and put heavy emphasis on the need to recognize the importance of the intensity of emotion from mild, through strong, to disorganizing. Arnold (1960) also argues convincingly for a separation of feelings and emotion. The search for clearly differentiating physiological patterns of arousal for each emotion has tended to confuse the issue. It seems clear now, from the work of Schachter (1962) and others, that essentially identical biological arousal can lead to either euphoria or anger and probably any number of other emotions.

If one examines feelings and emotion from the point of view of my first hypothesis, that feelings and emotion become understandable only in relation to self concept, it is possible to make meaningful distinctions. In these terms feelings arise as a result of a comparison between the incoming data and the self concept. If the data are irrelevant to worth, coping, expressing, or autonomy, the reaction is neutral. The individual is not interested or is even bored. If the data are relevant but also somewhat inconsistent with the self concept, then either a pleasant or an unpleasant feeling is experienced. If the inconsistency is in the direction of telling a person that he is more adequate than he had perceived himself to be, the feeling is pleasant. If the data indicate that he is less adequate, an unpleasant feeling arises.

The comparison process takes place at a level of functioning below awareness just as neural functioning does generally. As a result, we are aware of the feelings without consciously making the comparison. Thus, feelings are an individual's personal measure of the satisfyingness of data inputs to awareness. Feelings are essentially varying strengths of pleasantness and unpleasantness.

The values which we incorporate and live by are the ways of behaving which make us feel more adequate, the behavior which brings pleasant feelings. The value of honesty is much proclaimed but one which few people incorporate in its pure form. Most of us are relatively honest, but as a way of behaving this state is much diluted by the unpleasant experiences we have had at times when we were honest. If we are really concerned with the values which children incorporate, we must encourage them to express the feelings which the behavior we expect engenders in them. It may surprise or shock us when we find the amount of unpleasant feeling children experience in our school programs. There is no clue, however, to the effectiveness of our teaching which is more sure. The child who hates arithmetic may do what the school requires because it is the lesser of two evils, but we can be sure that he will not behave arithmetically when he is outside the classroom.

If we encourage children to share their feelings with us about the things which are happening in the classroom, this not only gives us a clue as to whether or not the learning is taking place, it also provides the opportunity to discover why the learning is unpleasant. Arithmetic is not inherently unpleasant for any children. Learning arithmetic becomes unpleasant only when it does not make sense to the pupil, or when wrong answers lead to embar-

rassment, or when the child is judged inadequate because he does not perform as rapidly or accurately as someone else. If a teacher is really concerned with the question of what Johnny will do outside the classroom rather than with the question of what he can do when coerced in the classroom, he will be constantly asking the question, "How do you *feel* about it, Johnny?" "How do you feel about the story—the music—the theme you wrote?" It is the feelings about things which determine what gets incorporated into future behavior.

In contrast to feelings, emotion, which is a much stronger bodily reaction, arises when the inputs from the outside world are widely discrepant from the perceived self and some outside situation or person is assessed as the source which is impeding or enhancing the self. Emotion is directed toward this situation or person and arouses the organism to action. When the self is seen as an object, the individual acts as though his self were an outside person and directs the emotion at himself. Emotion directed at the self may take many forms but is often experienced as depression, the emotion which accompanies the inability of the self to act when action seems to be called for.

The fact that an emotion calls for action takes on an added significance in our culture, which disapproves of strong emotion and its display. Were an individual able to hug somebody when he felt joy or strike somebody physically or verbally when he felt anger, the discrepancy between the self concept and the input could be resolved. Such action uses the energy that has been mobilized by the emotion; and as the energy is dissipated the people involved can explore new grounds for understanding.

I am not advocating direct physical action as a response to every emotion, but I certainly support the direct facing of anger, grief, love, or any other emotion by both parties in a situation evoking emotion. In most cases, if a teacher will merely listen sympathetically to a child's verbal expressions of emotion, the need for further action will disappear.

The experience of having one's emotional reactions appraised as unacceptable while one is growing up leads to a concept of self which is stunted and immature in the area of the ability to express one's self. The child perceives himself as an emotional person, but the models from which he gains his concept of adequacy build a picture which seems to value emotional control as the ideal. The large discrepancy between these two motivates the child to find

behaviors which he sees as more adequate. He tries to control or at least not to express his emotion, and verbally he denies to others that he is experiencing it.

If the child's model for adequacy was a mother or father or teacher who could be angry and could let the child be angry, he would in time learn to express emotion in ways that did not damage other people. I believe that the damage that some people do when they act on emotion is in large measure due to their inability to express their emotions appropriately. A picture of adequacy which stresses only control tends to suppress emotion until it becomes so strong that it breaks through the carefully constructed dam of control. At that point the emotion is in the area that Prescott calls disorganizing emotion. It produces disorganized behavior.

A similar deduction from self-concept theory can be made with regard to feelings. Not only may a child in our culture build a concept of adequacy based on a control model, but he frequently experiences inputs from his parents and teachers which tell him his feelings are wrong. He is told that "spinach is good," that "arithmetic is really lots of fun," that "this little needle prick isn't going to hurt one bit," and so on. The "good" boy or girl is the one who ignores his feelings and does as he or she is told. This leads people to distrust their feelings and reject their most important valuing process.

Such learning experiences are amazingly effective. I find that many college students with whom I work in a sensitivity training group have almost lost the ability to feel. When they are asked how they feel about being told that they impress others as being cold and manipulative, they answer, "Why, I don't know, I guess I don't really feel anything about it." When feelings and emotion begin to revive there is often a curious time delay. A person who is told on Monday that he interrupts others too much discovers on Wednesday that he was really very angry when he was told that, but he did not realize it at the time. This time lag gradually disappears as individuals turn from intellectualizing about things and begin to pay attention to what is happening inside their skins.

Feelings or emotion which are denied or unrecognized still affect behavior. This occurs when the feeling or emotion is translated into acceptable intellectual symbols. For example, when a white person comes into contact with a black person and pulls away, he might explain his behavior to himself or others as due to the fact that "you just can't reason with these people," and he

would avoid anyone like that. The explanation need not be true but it must be such that it is consistent with his self concept and avoids attributing negative qualities to himself. Prejudice, within this theory, is a product of conflicting elements in the self concept. He might have built a concept of adequacy in the area of worth which includes the idea that white is superior to black or that other attributes, such as speaking without an accent, dressing neatly, or cutting one's hair a certain length, are all connected with worth.

At the same time, being unprejudiced is also associated with worth. With such competing feelings, an individual will act in response to the unpleasant feeling evoked by the perceived negative attributes of another, but will verbally deny having had the unpleasant feelings because he wants to enjoy the pleasant feeling of being unprejudiced. If such conflicting feelings are a part of one's concept of adequacy, then one need never have learned prejudice against a particular race or group to still act in a prejudiced way toward anyone who is different from one's self concept. Hebb (1966, p. 245), after reviewing a number of studies, makes the statement, "An essential component in prejudice is the emotional reaction of human beings to the strange, to what is the same and yet different, to the thing that can cause a conflict of ideas." I would add the phrase, "or a conflict of feelings," to Hebb's statement.

Within this theory there are a number of possible causes for behavior that is ineffective or widely discrepant from some norm. The perceived self may be built on invalid appraisals, such as the case in which a father who really wanted a boy treats his daughter in ways which distort her feelings of worth and her ability to cope adequately as a girl. Behavior may also be ineffective when parental models from which the concept of adequacy is built are themselves markedly different from the dominant culture. This is probably the explanation for the lack of motivation in school of ghetto children who do not see learning to read or refraining from fighting as ways of becoming more adequate. Actually, they are likely to feel the exact opposite.

However, the point I would like to make in this paper is that one of the least understood sources of distorted and ineffective behavior is the suppression of the normal development of feelings and emotion. I have described how feelings operate as a personal measure of how well one's behavior is functioning in the service of self. Feelings are thus a guide to continuing or modifying behavior.

Emotion identifies for an individual the persons or situations which appear to him to be blocking or facilitating effective functioning and which direct him toward appropriate actions. The fact that our culture tends to suppress the "normal" functioning of both feelings and emotion robs people of an irreplaceable guide to effective functioning. This suppression of affective function also encourages the development of internal conflicts that lead to discrepancies between actions and verbal behavior, as is illustrated by denials of prejudiced behavior. Things may appear to go more smoothly in a society of controlled emotions, but the price we must pay is the stunting of emotional maturity.

The Promotion of Affective Development

At the risk of offending people in the teaching profession, I believe we must start with the assumption that each of us has in some degree suffered from the distortions and stunting of emotional maturity which I have described. It seems clear that if we are to foster effective development in the schools we must start with the teachers themselves. It has been suggested by Jersild (1955) and others that teachers should be psychoanalysed. This seems impractical when one considers the large number of teachers and small number of analysts.

There is now a newer technique available which is far less expensive and may move more directly to the target of teacher change. This is what is referred to as the T-Group or sensitivity training group. In this approach a group of eight to twelve people working with a trained leader spend all day together for a period of one to two weeks. To be most effective the group lives together in some place isolated from their normal contacts and from any temptation to sneak in a little work at home.

It is beyond the scope of this paper to describe the workings of such groups, but perhaps I can give you some feel for it by stating some ground rules by which such groups function. First, heavy stress is placed on responding to the here and now. The raw data with which members of the group work are their own experiences in this group. Second, members are encouraged to express their feelings and their spontaneous thoughts. Ponderous intellectualizing is taboo. Third, people are encouraged to use the personal pronoun, "I," rather than to talk about "they" or "some people think." We need to recognize and take possession of our

own feelings and thoughts. Fourth, communication is person to person as opposed to teacher to child or high status person to lower status person. Fifth, group members are encouraged to be experimental, to try out feelings and ideas which they might normally inhibit in their home situations.

This approach is not a panacea but, for most people, experiencing this way of being releases potential which they were not aware they had. They feel able to be themselves. The suppression of feelings and emotion is lifted. Life seems so much more real and important. It is a powerful experience that loosens the bonds of routine patterns of living.

If this sensitizing experience can be related to teaching, and there is some evidence that it can be (Gage, 1963, pp. 283-86), then I would predict that teacher behavior in the classroom would change in some of the ways indicated in the following paragraphs.

Teachers would see children differently and their expectations for them would change. The nature of this change is difficult to define, but it would show up in a warmer, more individualized, and more vital climate in the classroom. An interesting experiment is reported in the April 1968 *Scientific American* by Rosenthal and Jacobson (1968). In an elementary school heavily populated with "disadvantaged children," all children were tested with a new intelligence test which was unfamiliar to the teachers. The experimenters made a random selection of about ten children in each classroom and assigned them randomly into either a control group or an experimental group. The teachers had no knowledge of the experiment. The experimental treatment consisted solely of making a rather casual comment to each of the teachers that, in case they were interested, some children, who were then named, did well on the test. Their scores indicated that they would probably make unusual intellectual gains in the coming year.

All children were tested again at the end of the academic year with the same test. The results indicated strongly that children from whom teachers expected greater intellectual gains actually made those gains. Other aspects of the study ruled out the possibility that the teacher had spent more time with these children and, interestingly, the more the experimental children gained, the more all the other children in the classroom gained. The experimenters concluded that the explanation of these gains lay in more subtle teacher behaviors such as tone of voice, facial expression, and possibly touch and posture. These were, seemingly, the behaviors

which communicated the expectations. Climate of the classroom is something about which we have much to learn, but it is basically the feelings created in the children by the teacher. The increased awareness of and concern for feelings created by sensitivity training should make important changes in classroom climate.

Another change in teacher behavior which I would predict is that the classrooms themselves would take on some of the aspects of a sensitivity group. Children would be encouraged to express their feelings as well as their intellectualizations. Use would be made of artistic media at all grade levels, and children would be encouraged to talk or write about the feelings evoked by music or finger painting or a trip to the museum. Teachers would not be frightened by strong emotion. They would have learned the value of expressing emotion openly, and they would have learned the importance of having someone stand by and communicate indirectly that it is normal and acceptable to feel emotion. Some things or people really do make you angry, or afraid, or happy, or sad. The encouragement of feelings and emotional expression in the classroom would provide not only outlets but also opportunities to learn how it is possible to express one's self without damaging the self or others. Teachers can learn ways of helping children repair the temporary hurts once the feelings are out. Suppressed emotion tends to fester and to leave long-lasting resentments and hostility.

Another change in teacher behavior which I would expect to evolve would be a much greater emphasis on personal valuing of experiences. The intellectual methods of evaluating would not be ignored, but in most areas of school work the most important learning for a child is not the value to society, but the value to him, the learner. It is his feelings which tell him this, not the judgments of history or of the experts. This process is the source of values and we cannot leave the development of values to chance.

Finally, I believe that teachers who become sensitive to their own feelings and emotion would be able to encourage children to explore the conflicts which they feel. This does not take expert knowledge; the key behavior is listening and responding with understanding.

It is my guess that children who experience such a teacher would themselves become more alive. They would maintain awareness of their feelings and emotion and they would come to understand what they mean to themselves and to others. This would

increase their self understanding and ability to make choices which were beneficial to their own development. Some of the conflicts between personal worth and ability to cope might be resolved. Finally, it is my guess that learning in such a teacher's classroom would lead to greater tolerance for frustration, for ambiguity, and for the differences among people.

I realize that much of what I have said is speculative, but I believe that the self-concept approach to understanding feeling and emotion removes some of the confusion that now exists in education. It is becoming increasingly possible to think in terms of a curriculum to develop emotional maturity which would be parallel to and equal in importance to a curriculum for intellectual maturity.

References

- Magda B. Arnold. *Emotion and Personality*. (2 Vols.) New York: Columbia University Press, 1960.
- Walcott H. Beatty and Rodney A. Clark. "A Self Concept Theory of Learning." *Monograph*, San Francisco State College, 1962. Reprinted in: Henry C. Lindgren. *Readings in Educational Psychology*. New York: John Wiley & Sons, Inc., 1968.
- Nathaniel L. Gage, editor. *Handbook of Research on Teaching*. Chicago: Rand McNally & Company, 1963. pp. 283-86.
- Donald O. Hebb. *A Textbook of Psychology*. Philadelphia: W. B. Saunders, 1966.
- Arthur T. Jersild. *When Teachers Face Themselves*. New York: Bureau of Publications, Teachers College, Columbia University, 1955.
- Daniel A. Prescott. *Emotion and the Educative Process*. Washington, D.C.: American Council on Education, 1938.
- R. Rosenthal and L. Jacobson. "Teacher Expectations for the Disadvantaged." *Scientific American* 218 (4): 19-23; 1968.
- S. Schachter and J. E. Singer. "Cognitive, Social and Physiological Determinants of Emotional State." *Psychological Review*, Vol. 69; No. 5; 1962.
- Paul T. Young. *Motivation and Emotion*. New York: John Wiley & Sons, Inc., 1961.

Section II

*An Inventory
of Measures
of Affective
Behavior*

An Inventory of Measures of Affective Behavior

DONALD J. DOWD
SARAH C. WEST

Attitude Scales

DONALD G. BARKER, Texas A and M University, College Station
77843

Attitudes Toward Riding the School Bus

This instrument seeks a measure of the extent to which an individual student or group of students perceives the experience of commuting to school by bus as pleasant and satisfying, as neutral, or as unpleasant and frustrating. Norms provided for the scale are based on administration of an unsigned questionnaire to high school students.

The scale consists of 24 statements, 12 positive and 12 negative. Students respond using a five-point scale which represents strongly disagree, mildly disagree, not sure, mildly agree, strongly agree. For positive statements, scoring ranges from one for strong disagreement to five for strong agreement. This procedure is reversed to score negative statements. Responses to all 24 items are summed.

The 50th centile ($N=300$) was 65.4; corrected reliability coefficients varied from .90 to .95.

The instrument is printed in *Psychology in the Schools*, July 1966.

RALPH BENTLEY and AVESNO M. REMPEL, Purdue University, Lafayette, Indiana

The Purdue Teacher Opinionnaire (PTO)

This instrument is designed to measure teacher morale. It yields a total score, indicating general level of morale, as well as subscores on 10 morale dimensions such as Teacher Rapport with Principal, Teacher Salary, and Community Pressures. The PTO has been used in many research studies and has found use in determining general morale level of teachers in various schools and school systems. It has been

used effectively with both elementary and secondary school teachers. Normative data include raw-to-stanine score conversion and charts for comparison of obtained scores with quartile scores of 3023 teachers.

Test-retest correlations of the 10 dimensions range from $+.62$ to $+.88$. Validity is based on peer judgments of "high," "middle," and "low" morale levels, significant at the .05 level. Further validity data are furnished by mean scores of those leaving teaching (1965-66) with their occupational status in 1967-68. No time limit is effective; however, most finish within 25 to 30 minutes. The PTO is available in two forms, A or B. Scoring may be requested at the University Book Store, 360 State Street, West Lafayette, Indiana. (Specimen Set—\$1.00; Form A or B—\$4.50 for 25 copies). Further information can be obtained from Ralph Bentley.

DR. PRATIBHA DEO, Punjab University, Sector 14, Chandigarh, India: Department of Education

An Attitude Scale for Punishment

This is a self-administered scale which measures attitudes toward punishment of school children. It is a Likert-type scale composed of four combined factors of punishment—types of punishment, agents of punishment, situation of punishment, place of punishment. The Scale has been used in studies at Punjab University and some normative data are available. It is used primarily with those of high school age.

Reliability and validity are currently being studied. Scoring on the 92-item scale is based on five-point scales. Although originally constructed in Hindu, English forms are available. For information write the author.

DR. PRATIBHA DEO, Punjab University, Sector 14, Chandigarh, India: Department of Education

An Attitude Scale for Ragging

This scale is designed to measure quantitatively people's attitudes toward ragging, a bad practice found in Indian educational colleges. It can be given to any population that has a knowledge of ragging and can read English. Normative data are available on a sample of 885 students, teachers, and parents from two colleges in India. The scale can be given to groups.

Reliability and validity are unavailable as of yet. Scoring is done by a combination of Thurstone technique and Likert technique, from which a total score is obtained. For further information write Dr. Pratibha Deo.

JOHN E. JORDAN, College of Education, Michigan State University,
East Lansing

Attitudes Toward Handicapping Conditions

The instrument has been tested in nine nations. For details, see: *Attitudes Toward Education and Physically Disabled Persons in Eleven Nations*, by John E. Jordan. Latin American Studies Center, Michigan State University, 1968. 312 pp.

FRED N. KERLINGER, Division of Behavioral Sciences, School of
Education, New York University, New York, N.Y. 10003

Education Scale VII (ES-VII)

This is a Likert-type scale which measures two broad dimensions of attitudes toward education, progressivism and traditionalism, for research purposes. It is appropriate for graduate students and "fairly well-educated adults." This scale and several others were used in a set of studies which investigated the relations between attitudes toward education and perceptions or judgments of desirable traits of teachers.

This scale (as well as ES-VI) has been found factorially valid and reasonably reliable. The authors recommend the use of ES-VII where measures of progressivism and traditionalism are desired unless very high reliability is mandatory. In the latter case the longer form, ES-VI, should be used.

Some items from ES-VII are:

1. Learning is essentially a process of increasing one's store of information about the various fields of education.
15. We should fit the curriculum to the child and not the child to the curriculum.
30. Learning is experimental; the child should be taught to test alternatives before accepting any of them.

Both instruments are available from the American Documentation Institute or from Kerlinger. There are no restrictions on the use of the instruments, but the author would like to see results of any studies which use the instruments.

DONALD K. PUMROY, Counseling Center, University of Maryland,
College Park

Maryland Parent Attitude Survey (MPAS)

The purpose of this instrument is to measure parent attitudes toward child rearing. Four scales measure four different types of par-

ents—Disciplinarian, Indulgent, Protective, and Rejecting. The MPAS is strictly a research instrument, but normative data on high school and college students and parents are available. The MPAS can be used with anyone high school age and older, although it is designed primarily for parents. Scoring keys as well as T-scores are available.

Test-retest reliability correlations range from $+0.62$ to $+0.73$ for a three month interval, and split-half correlations ranged from $+0.66$ to $+0.84$ for the four scales. Some validity studies have been completed. Being a paper and pencil instrument, the MPAS is relatively easy to administer. More information about the MPAS may be obtained from the author.

RAY TOBIASON, Assistant Superintendent for Instruction, Puyallup Public Schools, Puyallup, Washington

Dissatisfaction Magnitude Scale (DIMS)

This instrument seeks to measure teacher dissatisfaction by comparing how the teacher feels now with how he would have to feel to be satisfied. On each page of the 15-page scale is a different item, e.g., "School Calendar" or "My Present Educational Role," to be judged by marking a seven-point scale between each of 20 bipolar adjective pairs such as good-bad, rational-emotional, or formal-informal. The teacher marks each scale according to how he feels at the time and then marks each scale again according to how he would have to feel to be satisfied.

Data were analyzed in terms of various groupings such as age, sex, teaching level, and aspiration level. Factor analysis of the scales was employed to strengthen their diagnostic potential. Correlations between DIMS and three alternate dissatisfaction measuring instruments were cited in claiming validity for the DIMS.

BENJAMIN D. WRIGHT, Department of Education, University of Chicago, Chicago, Illinois

Teaching Attitudes Questionnaire, 1962

The questionnaire is composed of two types of rating scales. The first consists of 26 bipolar adjective pairs which are used to elicit the subject's feelings toward himself as a person and teacher and toward his mother, father, and best-liked teacher. The second type of scale consists of 18 bipolar phrase pairs used to elicit the teacher's conception of himself and his best-liked teacher as teachers (in relation to behaviors specific to the role of the teacher in the classroom). Each pair, in both the adjective and the phrase type scales, represents a six-step continuum.

The following are examples of items:

strange 0 0 familiar
 deep 0 0 shallow
 acts old 0 0 acts young
 knows if we are trying 0 0 doesn't know if we are trying

Scale positions were assigned values of -3, -2, -1, +1, +2, +3. Biographical information, which was not scored, was requested on the first and last pages of the test booklet, and 20 questions about childhood relationships were also included. This information was used to compute correlations between scores and biographical variables such as social class and religion.

Extensive statistical data were reported. Anyone is free to duplicate the questionnaire, giving credit in a footnote. The cost of reproducing it has in the past run about 25¢ per questionnaire.

LAWRENCE S. WRIGHTSMAN, George Peabody College, Box 512, Nashville, Tennessee 37203

Philosophies of Human Nature Scale

This scale is designed to measure beliefs about human nature. The 84-item Likert-type scale provides six scores, one each for the dimensions of beliefs about human nature—untrustworthiness, altruism vs. selfishness, strength of will and rationality, independence from group pressures, complexity vs. simplicity, and similarity vs. variability. Norms are available for undergraduates and various occupational groups. The scale is applicable to those 14 and older. It has been used in over 50 studies, a bibliography of which has been prepared.

Test-retest and split-half reliability measures are reported as adequate and some information is available on validity measures. The scale may be administered to groups, and machine scoring is available through IBM. Further information may be obtained from the author.

Creativity

ELEANOR H. BARBEROUSSE, 17500 McDade Court, Rockville, Maryland 20855

Pupil Creativity Concept Q-Sort

The pupil is given 50 cards, each with a statement that can be used to describe a person's concept of himself in terms of traits that are exhibited by creative people. He is to arrange the cards in nine piles, from most like himself to least like himself, with the following numbers of cards in each pile:

1 2 5 10 14 10 5 2 1

Some of the 50 statements are:

- 5. I am interested in what everyone else does.
- 10. I value myself highly and I value others as highly as myself.
- 15. I am more comfortable when I am with people than when alone.
- 20. I seldom engage in any activity that is not safe.
- 30. I am guided by what other people expect of me.

Contact the author for further information.

DAVID A. DENNY, State University College, Oneonta, New York 13820

Denny-Ives Creativity Test

This test, suggested only for research purposes at this time, provides via the dramatic arts a multimedia assessment of pupil creativity at the sixth-grade level. It has been used in two research studies and also by classroom teachers in assessing pupil creativity for their own purposes. Although the author feels that the test is also appropriate in its present form for fifth graders, normative data are available only for the sixth-grade classes involved in the research studies.

The test consists of a tape-recorded story, tape-recorded music, and slides of various materials. After presentation of the story, slides (or a chart) depicting materials such as a piece of wood, netting, or green velvet are shown, and students are asked to list possible uses for the material in a dramatic presentation of the story. After listening to the music ("La Mer" by Debussy, "Til Eulenspiegel's Merry Pranks" by Strauss), they are to list possible uses for the music in putting on a play.

In Part II of the Creativity Test the student is asked to write a short description of a scene from the story as he would present it in a play. He is allowed to utilize only the specified props and materials.

Scores for fluency and redefinition are derived from Part I by counting the total number of ideas and the unusual ideas listed for objects. Part II, which yields scores for originality and sensitivity, requires the judgment of raters and thus results in some subjectivity. Guidelines for rating are offered in an effort to minimize subjectivity. Administration of the entire test requires about 45 minutes.

The instrument is available from the author. There is a charge of \$5.00 for reproducing the tapes and slides, and users must secure permission to mimeograph the test manual and pupil response sheets. The author is interested in obtaining normative data for the test.

DR. PRATIBHA DEO, Punjab University, Sector 14, Chandigarh, India: Department of Education

Tests of Creativity

Tests of creativity are being developed in a doctoral research project at Punjab. They will be standardized for use with Indian children and college students. There are six subtests, some of which are paper-pencil tests and others performance tests. Details of administration are given in the test booklet. Scores are obtained for fluency, flexibility, originality, inquisitiveness, and persistence. Although ideas were borrowed from Guilford and Torrance, original tests are included in the battery. The tests are available in English.

The results of an item analysis are being used, and further statistical data are being developed.

The tests are not yet priced, but they will be available from the author at a reasonable price.

JACK R. FRYMIER, School of Education, The Ohio State University, Columbus 43210

Ohio State Picture Preference Scale (OSPPS)

The OSPPS provides a nonverbal measure of creativity, delinquency-proneness, and motivation toward school. It has been used with over 5,000 adolescents and young adults, and with some very young children. Using this instrument, efforts are now under way to predict college success among disadvantaged children. More than 10 studies utilizing the OSPPS have been made with medical students, prisoners, delinquents, underachievers, and overachievers. No normative data are yet available. Item analyses and correlations have been completed in several studies.

The OSPPS consists of 100 items, each having a pair of pictures which are similar in certain ways but fundamentally different in others.

The responses are categorically simple (making a choice of one of the two pictures in each pair). It is assumed that each time a respondent makes a choice, he brings his perceptual apparatus and his previous experience to bear on the decision involved and "projects" himself into the response, at least to some degree. Several experimental keys are available from the author for research uses.

The OSPPS is not for sale yet but may be borrowed from the author.

PAUL McREYNOLDS and MARY ACKER, Behavioral Research Laboratory, V. D. Hospital, Palo Alto, California

Obscure Figures Test

The OFT is a potentially useful instrument in studies concerned with creativity, curiosity, and reactions to novel stimuli. It has been used in several studies with both normal and psychiatric patients.

Form I of the OFT is comprised of 40 line figures which can be perceived as representing various objects. The respondent's task is to think of something that each figure might represent. In addition, 40 figures which can be utilized as an alternate form of the test are available, but little work has been done with them. The test may be administered individually or to groups. *Working time is usually limited to 10 minutes*, though many do complete the test in less time.

Each response is awarded from zero to three points, and the total score is the sum of the points for the 40 figures. Guidelines are provided for scoring each item in an effort to reduce subjectivity of scoring.

Normative data for a total of 410 normal cases are reported in the manual. The instrument is reasonably reliable for assessing cognitive innovation. While adequate evaluations of its validity are not yet available, the test appears to possess sufficient face validity to justify further research along this line. Correlations significant at the .05 level have been reported between the OFT and the Plot Titles Test and between the OFT and OPI Originality Scale.

The OFT is not for sale but will be *loaned* to appropriate persons on request. Contact the author for further information.

THOMAS J. ROOKEY, Division of Research Design, Bureau of Research, Administration, and Coordination, Department of Public Instruction, Harrisburg, Pennsylvania

Pennsylvania Assessment of Creative Tendency

Designed as a test of potential rather than practice, this instrument assesses the factors which underlie a student's creative behavior. It is

based, with certain reservations, upon the work of Torrance and measures nine traits—self-direction, evaluative ability, flexible thinking, original thinking, elaborative thinking, willingness to take risks, ease with complexity, curiosity, and fluent thinking ability.

This self-report instrument consists of 46 Likert-type items. It has been administered to over 300 fifth-grade students in the Harrisburg, Pennsylvania, area. Scores correlated well with subjective teacher ratings of pupils.

The PACT was designed for the Pennsylvania Quality Assessment of Education and will be more extensively used in the future.

Contact Rookey for further information.

E. PAUL TORRANCE, Department of Educational Psychology, College of Education, University of Georgia, Athens 30601

Torrance Tests of Creative Thinking

These tests assess various kinds of creative functioning, various types of creative development, and outcomes of experimental materials and methods. They are also useful in identifying certain types of creative potentialities. They have been used in approximately 200 research reports, theses, and dissertations over a period of nine years but have received limited clinical use.

Normative data for each grade from first through twelfth are reported in the technical-norms manual. The tests are appropriate for all ages beginning with four year olds and for all occupational groups. Statistical data are also reported in the manual.

Individual administration is necessary with kindergartners through third graders, except with the figural tests. All other tests may be group administered.

Detailed scoring guides are available for all forms. There is a scoring service which is offered by the publisher, Personnel Press, Inc. A description of some of the new types of items for measuring the creative thinking abilities is also available. A specimen set (including four forms, all four scoring guides, scoring worksheets, and the technical-norms manual) costs \$3.00 and is available from: Personnel Press, Inc., 20 Nassau Street, Princeton, New Jersey 08540.

Interaction

JOHN D. ALCORN, Box 478, Southern Station, Hattiesburg, Mississippi 39401

The Interpersonal Orientation Scale (IOS)

This self-administering scale was designed to assess (a) interpersonal relatedness on an altruistic-manipulative axis and (b) preference levels for five manipulative techniques including coercing, evaluating, masking, coaxing, and postponing. It has been used with school counselors, teachers, and administrators; social workers; finance company employees; prison inmates; and recent graduates of educational counselor programs. The authors (Alcorn and Dr. Everett D. Erb, East Texas State University, Commerce, Texas 75429) are currently preparing percentile norms. The IOS is appropriate for study of any category of behavior involving interpersonal relationships and especially helping relationships.

The IOS consists of two sections. The first section with 52 items poses situations and asks the respondent to identify the action he feels is most appropriate. The following is a sample taken from Section I:

Situation #2. There is a man in your community who has a great deal of ability but demonstrates little ambition toward making an adequate living for his family. If you were his wife, would you

2. a. Simply stand behind him and provide moral support.
b. Make sure he is aware of his family's plight due to his lack of ambition.
3. a. Try to show understanding for his feelings.
b. Point out his responsibilities toward himself and his family.

Section II simply requires the respondent to register agreement or disagreement with an assortment of statements such as the following:

15. Things done quickly are usually half-done.
25. When in trouble it is better to keep your mouth closed.
42. I have no use for sissies.

From 30-35 minutes are required to complete all six subscales. Answer sheets are scored using six overlays.

Researchers may obtain permission from the author to reproduce their own copies of this instrument.

EDMUND AMIDON, Temple University, Philadelphia, Pennsylvania

The Verbal Interaction Category System (VICS)

The VICS, based upon the Flanders System of coding verbal behavior every three seconds, is a technique designed to analyze classroom

interaction so that teachers can identify their own styles of verbal behavior and consciously vary that behavior according to what they wish to accomplish.

There are 12 categories which must be memorized by the observer. Categories are tallied every three seconds in sequence in a column. If the verbal behavior changes before the three-second interval ends, this change is always recorded. Ultimately, the categories are entered in a 17-row by 17-column matrix which presents information about the type, sequence, and amount of verbal behavior. This technique has been used in many research projects and teacher-training programs; numerous pertinent articles exist in professional journals. The VICS is appropriate for teachers and student teachers as well as for administrators and supervisors. The author cites the copious research data from various studies as evidence of the validity of the technique.

Further information is available in a book by Amidon and Hunter, *Improving Teaching: The Analysis of Classroom Interaction*. New York: Holt, Rinehart and Winston, Inc., 1966.

DAVID N. ASPY, 2027 N.W. 7th Lane, Gainesville, Florida

Truax Scales for Empathy, Congruence, and Positive Regard

This is a technique designed to assess the helper's (counselor, teacher, parent, etc.) levels of Empathy, Congruence, and Positive Regard in verbal interaction with the client or student. A recording is made of interaction and then these recordings are rated according to preestablished scales of the above. These scales have been used in studies of student achievement and the levels of Empathy, Congruence, and Positive Regard of the teachers. Carkhuff, Truax, and Rogers have used this type of scale in a psychotherapeutic context.

It is reported that training procedures have developed raters whose intrajudge reliabilities are consistently above +.90 and interjudge reliabilities are above +.50. The scales used are composed of five levels, with level one representing the lowest level of interpersonal functioning. Additional description of the scales can be found in recent issues of the *Journal of Counseling Psychology* or by contacting David Aspy (see above).

ELEANOR H. BARBEROUSSE, 17500 McDade Court, Rockville, Maryland 20855

Sociometric Reputation Nomination Scale

This is a scale used to assess peer group status as related to: friends, stars, isolates, intellectuals, leaders, teacher's preference, extro-

verts, introverts, creatives (+ & -), enemies, and humorists. It was developed in conjunction with the author's dissertation. Normative data were compiled on the eighth-grade populations of two Alabama towns. It can be utilized in sixth through tenth grades.

As the scale is used mainly for description, reliability and validity are not reported. However, the instrument compared favorably with background data on the normative sample. The scale requires the student to write three names for each of the categories in response to stimulus questions. Scoring is rather laborious. Further questions could be addressed to the author.

JAMES DUNCAN and JACK R. FRYMIER, The Ohio State University, Columbus 43210

Duncan Teaching Situation Reaction Test (TSRT)

The purpose of the TSRT is to measure a person's perceptions of children and youth as well as his reaction to teaching situations. John Hough expanded the original 36-item test by adding an additional 12 items, which were designed to assess open- or closed-mindedness. Considerable research has been accomplished on its validity and reliability.

The TSRT consists of a case example designed to measure teachers' abilities to work through some of the problems of handling a classroom group. Teachers are given certain information about the classroom group and the working situation; then they are presented with several questions. Each question has four alternative responses which teachers must rank as their first, second, third, and fourth choices. This procedure is repeated through a series of problem situations. The case studies were designed so that teachers could respond regardless of their subject fields. In all there are 11 problem situations and 48 questions.

For further information contact the author.

NED A. FLANDERS, College of Education, University of Michigan, Ann Arbor 48104

Interaction Analysis

This is a technique to code verbal communication in the classroom using standardized categories and recording procedures. It has been used in basic research on teaching behavior, in preservice preparation of teachers, and in-service training of teachers. Observers must be trained in the scoring and analysis procedures, which are quite complicated.

For detailed explanation of the system refer to: Aidon and Hough. *Interaction Analysis: Theory, Research and Application*. Reading, Massachusetts: Addison-Wesley Publishing Company, Inc., 1967.

FRED K. HONIGMAN, Franklin Institute Research Laboratories, 20th and Race Streets, Philadelphia, Pennsylvania 19103

Multidimensional Analysis of Classroom Interaction (MACI)

MACI, a system for categorizing teacher and pupil behaviors, was designed to serve as both a conceptual model for analyzing teaching and an observational instrument for coding and quantifying the behaviors of teachers and pupils in the classroom.

MACI focuses on three aspects of the teacher-pupil relationship—the affective, the cognitive, and the procedural. The recording procedure, writing a symbol every three seconds to represent the behavior taking place, was borrowed from Flanders. However, several MACI observational innovations expand the amount of information communicated by the behavior categories. Another difference between the MACI and Flanders techniques is that in MACI nonverbal behaviors are handled the same way as verbal behaviors and are accommodated within each of the regular categories.

Like the Flanders System, MACI uses a matrix as a means of organizing system-structured data. MACI data are analyzed on a dimension-by-dimension basis in terms of a formal, programmed, interpretive structure which utilizes three kinds of data—the frequency of occurrence of each behavior or event, the frequency with which different sequences of events occurred, and the typical length of performance of each behavior or event.

A variety of computer programs are available in FORTRAN IV for either research or teacher feedback.

Inter-observer reliability will be available soon. Validity studies are described in the book, *Multidimensional Analysis of Classroom Interaction*. Villanova, Pennsylvania: The Villanova University Press, 1967.

PAUL G. LIBERTY, Southwestern Cooperative Educational Laboratory, 117 Richmond, N.E., Albuquerque, New Mexico 87112

SWCEL Classroom Observer Rating Schedule

This instrument assesses classroom atmosphere in terms of both children's behavior and teacher performance. Pupil's behavior is interpreted according to *Taxonomy of Educational Objectives, Affective Domain*. There are three subsections which focus respectively on receiving, responding, and valuing. The instrument has been used in studies of the effects of various reinforcement strategies. Its use is most appropriate in the lower grades.

Satisfactory inter-rater reliability was reported. The author stated that clearer definition of categories and possible regrouping of items are

being considered. However, he considers the instrument satisfactory for exploratory research.

The observer records the number of pupils who are or are not exhibiting the stated behavior. He observes for a total of 20 minutes, spending specified periods of time (two minutes, three minutes, 15 seconds, etc.) on the subsections.

Copies of this rating schedule may be requested from the author.

DONALD M. MEDLEY, Educational Testing Service, Princeton, New Jersey 08540

Observation Schedule and Record—OScAR 2a

OScAR 2a was devised to measure three dimensions of classroom behavior on the basis of direct observation in the classroom. A factor analysis based on the initial use of this technique yielded three orthogonal factors called Emotional Climate, Social Organization, and Verbal Emphasis. It is appropriate for use in elementary classes, but there are no normative data available. The three factor scales have reliabilities of about .80 (based on 12 half-hour observations).

The system is relatively easy to learn, but some training is a necessity. Although the OScAR 2a is unpublished, a description is found in the *Journal of Educational Psychology* 49: 86; 1958.

DONALD M. MEDLEY, Educational Testing Service, Princeton, New Jersey 08540

Observation Schedule and Record 5V (OScAR 5V)

It is the purpose of this technique to measure the learning environment in the classroom by coding verbal behaviors of the teacher as observed either in the classroom or on video tape. The OScAR 5V has been used in several studies, including studies of changes in teacher behavior with experience and follow-up studies of the graduates of teacher-education programs. It is appropriate for both elementary and secondary classrooms.

No normative or statistical data are available. Earlier versions have shown reliabilities from .40 to .80, depending on the size of behavior sample coded.

Observers must be trained in the system. There is a system for interpreting, coding, and scoring behaviors. Contact the author for further information.

LAWRENCE A. PERVIN, Department of Psychology, Princeton University, Princeton, New Jersey

Transactional Analysis of Personality and Environment (TAPE)

TAPE provides an assessment of student-college interaction through student perceptions of the college and/or analysis of the college as a total system. The instrument has been used in 30 colleges around the country, from which the normative data (means and S.D.'s on concepts of College, Self, Students, Faculty, Administration, Ideal College) were derived. TAPE utilizes the semantic differential technique. It is appropriate for use with college students.

Reliability measures (product-moment) have provided coefficients ranging from +.40 to +.99 for the concepts of College, Self, Students. For further information contact the above address.

The following are scale factors and sample scales derived from a three-mode factor analysis.

<i>Factor</i>	<i>Sample Scales</i>
Impulsivity-inhibition	sober-intoxicated
Goal-directed activity	disciplined-undisciplined
	motivated-undirected
Sensitivity	industrious-tranquil
	feminine-masculine
	sensitive-insensitive

In all, 13 factors were so derived. A reprint of an article from the *Journal of Educational Psychology* contains further explanation of the instrument.

ANITA SIMON, Temple University, Ritter Hall—Room 263, Philadelphia, Pennsylvania, and YVONNE AGAZARIAN, Hall Mercer Community Mental Health Center, Pennsylvania Hospital, 8th and Spruce Streets, Philadelphia, Pennsylvania

Sequential Analysis of Verbal Interaction (SAVI)

This technique analyzes the verbal behaviors of a group in order to describe how the group is handling its communication to determine interpersonal difficulties and strengths, and to prescribe alternative patterns of behavior based on diagnosis of SAVI data. It can be used with any group and in fact has been used in several studies involving a variety of kinds of groups.

Using this technique, a highly trained (20-40 hours) observer records a code letter every three seconds to represent the kind of behavior

occurring. The coded tallies are entered into a matrix which reveals every sequence pair of behavior. The matrix reveals what kinds of behavior brought what kinds of responses, what kinds of behaviors the group is using, and what the group's mode of working together is.

Reliability of about .90 has been reported.

SAVI is particularly appropriate as a training tool for people in the role of change agent, e.g., teachers, administrators, supervisors, counselors. The use of SAVI provides an opportunity for people to obtain objective feedback about their styles of interacting and clues to new kinds of behaviors which might be helpful in their work. "Sequential Analysis of Verbal Interaction" is part of a compendium for classroom observation systems: *Mirrors for Behaviors*. Anita Simon and E. Gil Boyer, editors. Philadelphia: Research for Better Schools, Inc., 121 South Broad St., 1967.

Further information is available from the author.

ROBERT L. SPAULDING, 1516 Woodburn Road, Durham, North Carolina 27705

Coping Analysis Schedule for Educational Settings (CASES)

CASES measures child behavior in any social setting where societal or adult expectations provide the structure for evaluation, especially in educational settings where specific goals regarding desirable and inappropriate child behavior are set. It is currently being used both for teacher-training and for measurement of pupil and teacher behavior change.

Norms have been obtained for approximately 130 children of ages two through nine years. Means and standard deviations by classroom activity, age, sex, and race are available for a relatively restricted sample drawn from lower-middle and low income families. Reliabilities of observations and recordings have been satisfactory. Concurrent validity is being established through a variety of studies.

Observers employ either time samplings or continuous behavior recording, utilizing a tally sheet, signal generator, and clipboard. Data from tally sheets are analyzed in terms of percentages of frequencies.

Case studies employing CASES may be obtained from Mrs. Joan First, Information Director, Education Improvement Program, North Carolina Mutual Building, Mutual Plaza, Durham, North Carolina 27701.

CASES is available for \$1.00. Contact Robert L. Spaulding, Director, Education Improvement Program, 2010 Campus Drive, Duke University, Durham, North Carolina 27706

HENRY H. WIESEN, 1740 Terrace View Drive, West Columbia, South Carolina 29169

Organizational Climate in the Classroom (OCIC)

It is the purpose of this instrument to determine the degree and type of teacher control as perceived by the students in a classroom. The instrument was used in the author's dissertation and by several elementary school principals as a screening technique for prospective teachers. It is appropriate for use with students 10 or more years old.

The test consists of 12 multiple choice items which relate to the students' classroom behavior. The stem of each item is "My classmates and I . . .," and three statements which respectively represent loosely, moderately, and rigidly controlled classrooms are provided. The student is to select the one statement that most accurately describes his classroom. Three points are scored for each loosely controlled statement, two for each moderately controlled statement, and one for each rigidly controlled statement.

The test-retest correlation coefficient was .99, and the teacher-pupil test score correlations were .97 and .90.

The instrument may be duplicated and utilized.

E. N. WRIGHT, Toronto Board of Education, 155 College Street, Toronto 2B, Ontario

Draw-A-Classroom Test

The Draw-A-Classroom Test was designed to reveal the world the child perceives and how this world is influenced by school experiences. It also affords information about the developing concepts and ideas of the child in his mental, emotional, and social areas of growth.

The test is administered to pupils four to ten years of age by their classroom teachers. Children are given paper, crayons, and standard instructions, "Look all around the room and draw your classroom." No time limit is imposed. When finished, each child is asked to tell his teacher about his drawing, and his words are recorded on the face of the drawing.

A coding system has been devised in order to compare drawings of the same child from year to year and to compare the drawings of different children of the same age. Dealing with space, people, and objects, 83 categories were identified giving an average total of over .80 inter-rater reliability. At the time of the report (September 1966) the results had not been analyzed, and normative data were not available.

For research purposes, scoring categories and the manual of instructions may be obtained from the Research Department of the Board of Education for the City of Toronto.

Miscellaneous

HELEN S. ASTIN, 950 Lathrop Place, Stanford, California

Situational Test of Empathy

This is a test designed to measure empathetic ability. The STE contains 10 client statements taped by a professional actor. They represent different types of clients with different types of problems. The subject has to respond verbally to the statements as though he is in a counseling situation. It can be used with most groups concerned with the selection of persons for counseling or clinical type work or with any situation that requires empathy.

Intrajudge reliability of responses, ranked by professional counselors for degree of empathy, was $+ .82$.

Further information and copies of the tapes are available from Helen Astin.

ROBERT E. BILLS, University of Alabama: College of Education

Index of Adjustment and Values (IAV)

The IAV was designed to measure variables (self concept, self-acceptance, concept of ideal self, discrepancy between self concept and ideal self, and perceptions of how other people accept themselves) of importance to client-centered therapists and perceptual theorists. Administered individually or in groups, the IAV reports three scores. High school and college norms are available. Two forms, one for high school seniors and adults, the other for elementary, junior high, and senior high make the IAV widely useful.

Reliability measures report split-half coefficients ranging from $+ .53$ to $+ .91$ and test-retest (six weeks) coefficients from $+ .83$ to $+ .92$. Extensive validation procedures have been applied, and concurrent and construct validities established. Further information may be requested from the author.

CHARLES F. COMBS, Department of Counseling and Educational Psychology, Arizona State University, Tempe, Arizona 85281

Combs School Apperception Test

This instrument provides a measure of perceptual understanding in elementary or junior high school children. It has been used in many research studies and normative data are available.

For details of the instrument contact the author. Price: \$15.00.

DR. PRATIBHA DEO, Punjab University, Sector 14, Chandigarh, India: Department of Education

A Picture Test for Social Distance

This test was developed to measure the social distance between Indian people and (a) Pakistanis and (b) Chinese, both neighbors of India. It has been utilized in a few studies of social distance, and norms are available on 200 university students in India. It is appropriate for almost any population. The subject views pictures appropriate to the populations involved, and his response is then recorded by the tester as plus, zero, or negative.

Test or retest reliability is reported as $+.89$. Validity was determined through cross validation procedures with the Bogardus Social Distance Scale, giving coefficients of $+.69$ with Pakistanis and $+.72$ for the Chinese. Although primarily an individually administered test, it can be adapted to groups. Further inquiries should be addressed to the author.

DR. PRATIBHA DEO, Punjab University, Sector 14, Chandigarh, India: Department of Education

D-I Inventory

This inventory represents an attempt objectively to identify disciplined and undisciplined students through the use of a forced choice projection technique. It has been used in several studies in India, and norms based on 850 students from various Indian colleges are available. It can be used with high school, college, and university students.

Test-retest reliability (45 days) is reported at $+.68$; validity measures, which ranged from $.80$ to $.89$, were obtained through point biserial correlations with teachers' ratings. The inventory can be administered to groups. Further information may be obtained from the author.

WILLIAM R. FISHBURN, Department of Counseling and Guidance, School of Education, Indiana University, Bloomington 47401

Group Counseling Evaluation Scale (Form II)

This 50-item scale measures the perceptions of group members who participate in group counseling. It has been used experimentally only with approximately 100 subjects. A self-administering scale, it requires about 20 minutes to complete. Responses are scored on a four-point scale of strongly agree, agree, disagree, and strongly disagree. The following are sample items from the scale:

1. Confidentiality was maintained by the group members.

- 6. The groups were structured with a known purpose.
- 29. It is important for me to know my needs, attitudes, values, beliefs.
- 36. I talked about my feelings in the group sessions.

The scale may be reproduced for research purposes if results are returned to the author.

DALE B. HARRIS, Department of Psychology, Pennsylvania State University, 117 Burrowes Bldg., University Park, Pennsylvania

Goodenough-Harris Drawing Test

This test is designed to provide the user with a conveniently administered, intrinsically interesting measure of children's intellectual maturity. It is one of the most widely used tests for children in clinical settings and has been used extensively in cross-cultural research. The test is appropriate for children ages three to fourteen or fifteen and with mentally retarded adults and adolescents.

Further technical information regarding the test can be found in: Dale B. Harris. *Children's Drawings as Measures of Intellectual Maturity*. New York: Harcourt, Brace and World, Inc., 1963, or by contacting the author at the above address.

ROBERT B. HAYES, Director, Bureau of Research, Administration, and Coordination, Pennsylvania Department of Public Instruction, Box 911, Harrisburg, Pennsylvania 17126

Hayes Pupil-Teacher Reaction Scale

The purpose of this scale is to measure attitudes of pupils toward the teaching of their teachers. It has been used in two separate U.S. Office of Education projects which involved tenth- and sixth-grade students respectively. Normative data are available from both studies.

The scale consists of 20 items, nine of which are answered on a four-point scale and the remainder on an agree-disagree basis. The following are items from the scale:

- 2. This teacher really causes you to think.
 - a. Most of the time
 - b. Often
 - c. Sometimes
 - d. Seldom or never
- 7. His instruction is
 - a. Extremely challenging
 - b. Very challenging
 - c. Somewhat challenging
 - d. Not very challenging or usually unchallenging

11. This instructor is one of the best.
 - a. Agree
 - b. Disagree
15. His lessons are poor.
 - a. Agree
 - b. Disagree

Coefficients of consistency for sixth graders were .73 after three weeks and .58 after 21 weeks; for tenth graders they were .73 and .67.

SEYMOUR LEMESHOW, 15 Bedford Road, Kendall Park, New Jersey

Teacher Operational Problems Identification

A self-reporting, self-administering checklist, this instrument may be used for teacher counseling, in-service education, self-evaluation, and research by identifying operational problems encountered by secondary classroom teachers. Studies using the instrument have not been completed, but percentile norms are available.

Validity is claimed on the basis of the scope, source, and coverage of the problems. The test-retest coefficient of stability was reported to be .96. Median administration time is 30 minutes.

The instrument does not yield precise measurement scores; however, provision is made for weighting and counting scores of frequency and intensity. The following are directions and some sample items from the checklist:

If a statement represents a problem for you, do the following:

First: determine how often you encounter this problem.

often = 3; moderately = 2; infrequently = 1.

Second: determine how intensely you feel the problem.

severely = 3; moderately = 2; mildly = 1.

	Frequency			Intensity		
Students being unable to read	3	2	1	3	2	1
Having too much clerical work	3	2	1	3	2	1
Classes being too large	3	2	1	3	2	1

This instrument is being published by Educational Testing Service and will be available from the ETS Cooperative Test Division at a price to be established.

ROSS L. MOONEY, 278 East Longview Avenue, Columbus, Ohio 43202

Mooney Problem Checklist

The purpose of this instrument is to help students express their problems. It has been used in myriad studies and also in clinical and

counseling work. Junior high, high school, college, and adult levels are available.

The checklist is self-administering. Copies of the test may be obtained through: Test Division, The Psychological Corp., 304 East 45th Street, New York, New York. Much published data are available, and many published studies are to be found in the professional journals.

JACK NOVICK, Jewish Board of Guardians, New York, New York

Deviant Behavior Inventory (DBI)

The DBI consists of 237 items of deviant behavior covering such areas as the physical system, speech, thought, affects, self feeling, peer and adult relations, and lies. The DBI is administered in the form of a card sort, each item typed on individual cards. The respondent is then requested to judge on behavior occurring within the last six months and to sort the items as "True," "Not Sure," "False." Later an inquiry is made of the "True" and "Not Sure" responses.

The following are items from the DBI:

- 6. Has poor appetite: can't eat or just nibbles at food.
- 107. Blames himself even when he has done nothing wrong.
- 179. Has taken money that does not belong to him.

Further information and copies of the DBI can be obtained from the author.

JIN ONG, Linfield College, McMinnville, Oregon 97128

The Opposite-Form Procedure in Inventory Construction and Research

It is the purpose of this technique to develop the opposite-form procedure in constructing inventory-type tests. Until the author publishes this procedure, however, it will not be available for practical use.

This procedure is appropriate for use with junior high, senior high, college, and adult groups and has been used with adult extension students and college students. It assumes the form of an ordinary paper and pencil test and can be group-administered.

More information about the technique may be found in: "Reliability of Special Tests in Measuring Personality." *Psychological Reports* 19: 915-22; 1966. For a price of \$2.50, the instrument is available from Vantage Press, 120 West 31st Street, New York, New York 10001.

ERNEST L. PETERS, Director, Division of Cooperative Research Services, Department of Public Instruction, Room 214, Executive House, Harrisburg, Pennsylvania 17126

Pennsylvania Citizenship Assessment Instrument—Fifth Grade

The purpose of this self-report instrument is to assess the acquisition of habits and attitudes associated with responsible citizenship. Pilot studies involving 125 students have been completed, and use with 3,000 students is planned for April 1968. Item analysis and factor analysis with subsequent Varimax rotation were used to identify items to be included in the test package. Analysis of data and refinements of items will be made after the April testing; reliability and validity studies will run in the fall of 1968.

Composed of two parts, both of which utilize the Likert technique of scoring, the instrument questions the student about his behavior as well as his beliefs. In regard to the former, Part I consists of 20 statements to which the student responds—never, very seldom, sometimes, most of the time, or always. Questions of beliefs are found in Part II, where the student responds to 17 statements on a five-point scale which ranges from disagree strongly to agree strongly.

Anonymity should be assured in order to obtain valid and reliable responses. It may be necessary to administer items orally to poor readers.

This instrument and an eleventh-grade form of it are available for the cost of printing. For further information, contact the author.

ROBERT T. REEBACK, Southwestern Cooperative Educational Laboratory, 117 Richmond N.E., Albuquerque, New Mexico 87106

The Vigilance Game

This is a technique to assess and to provide the basis for controlling the level of attentiveness of pupils during a lesson. The Vigilance Game has been used by two teachers with five- and six-year-old Navajo children in classes in English as a second language.

The teachers sent signals to the children at semi-random intervals which required a response from them. The responses were recorded on video tape. This was used as a measure of attention and was positively related to performance in the primary task, the class lessons. A group-based reinforcement procedure was effective in increasing the level of attention.

ROBERT L. SPAULDING, 1516 Woodburn Road, Durham, North Carolina

The Spaulding Teacher Activity Rating Schedule (STRS)

This instrument is appropriate for assessing teacher behavior in terms of the techniques used by teachers to modify child behavior in motor, social, or cognitive areas. It is currently being used in the Durham Educational Improvement Program and has been used previously in teacher training. It employs trained observers who use either time sampling or continuous behavior recording.

Technicians trained in the technique report observer reliabilities in the order of $\pm .90$. Concurrent validity is now being established. Write: Mrs. Joan First, Education Improvement Program, North Carolina Mutual Building, Mutual Plaza, Durham, North Carolina, or write to the author for additional information.

LAWRENCE H. STEWART, Department of Education, University of California, Berkeley

Interest Assessment Scales

This instrument is an equisection scale that measures interest in eight areas. It has been used to differentiate well-defined criterion groups based on curriculum choices, vocational aspirations, and college major and to study the factor structure of interest. The scale is essentially non-normative and can be used with groups of students in the tenth grade and beyond. To date, the instrument has been utilized for research purposes only.

Stewart reported that the scales are fairly reliable over intervals of several months. Scoring can be accomplished by hand or IBM 7094. Descriptive materials of multivariate analysis of subjects are available. Further information on validity and scale development can be obtained from the author.

SUTTON-SMITH and B. G. ROSENBERG, Bowling Green State University, Bowling Green, Ohio

What I Like To Do (An Impulsivity Scale)

This instrument was submitted by Mary A. Barbour, who used it in conjunction with her doctoral dissertation. It consists of statements to which the subject responds either "true" or "false." The following are examples of statements in the scale:

I like to keep moving around.
 My home life is not always happy.
 I play hooky sometimes.
 I like throwing stones at targets.

Degree of impulsivity is indicated by the number of "true" responses; the child who circles more T's than F's is more impulsive than the child who circles more F's than T's. Reliability, as estimated by the test-retest method, is $+.85$.

Some statistical data for a sample of 342 sixth-grade pupils are reported in Barbour's dissertation.

WALTER L. THOMAS, 3860 Plainfield, N.E., Grand Rapids, Michigan 49505

Differential Value Profile

The purpose of this instrument is to identify and profile six value dispositions (aesthetic, humanitarian, intellectual, power, material, and religious) in most adolescents and adults in our society. The DVP provides information for studying the effects of various educational experiences on personal values, the prediction of attrition and dropouts, the guidance of vocational choices, the analysis of entering and graduating students, the prediction of grades in courses and programs, the guidance and forecast of educational choices, and the differentiation of students in different types of institutions.

Developed by factor analysis, the DVP has excellent reliability and established validity. Norms are available for large samples of high school and college students and noncollege adults. Weighted response mode, raw scores, converted scores, confidence bands, and standard errors of measurement are available. Modified Likert-type response mode is used. The DVP may be machine or hand scored, and statistical services are offered by the publisher.

The author asserts that the DVP has obtained higher predictive coefficients than the SAT or the ACT for college grades.

The instrument is available upon order from: Educational Service Co., 3860 Plainfield, N.E., Grand Rapids, Michigan 49505.

Prices: manual	\$4.00 each
test booklets	.15 each
answer sheets	.05 each
scoring keys	1.00 per set of six
scoring service	.45 per student

JOHANNA C. VAN LOOY, 24 Macaltioner Avenue, Woodstown, New Jersey 08098

Van Looy's Expectancy Scale

This scale was designed to measure pupils' self-expectations and their perceptions of their parents', teachers', and peers' expectations of them. The instrument has been used in several studies. It is appropriate for ten- to twelve-year-old children, and norms are available for upper-middle class students. Reliability is satisfactory. Two sources of external criteria were used as evidence of validity.

The scale can be administered to groups, and no time limit is imposed. It consists of 48 items such as the following:

I am expected:	by my parents	by my teacher	by my friends	by myself
1. To take care of my personal property	—	—	—	—
2. Not to fight	—	—	—	—
17. To be popular	—	—	—	—
42. To finish a job once I've started it	—	—	—	—

Students make four responses to each item according to a scale with five levels—never, sometimes, about half the time, most of the time, always.

Single copies of the instrument will be sent upon request. The author is interested in receiving results obtained with various samples.

Motivation

JOEL ARONOFF, Department of Psychology, Michigan State University, East Lansing

Sentence Completion Test

The purpose of this projective technique is to obtain the level of motivation within Maslow's theory of the hierarchy of needs. In principle it is applicable to all groups, although past usage has been limited to adult males and adolescent children on the West Indian island of St. Kitts. It may be administered individually or in groups. There is a scoring system which is explained in Aronoff's book, *Psychological Needs and Cultural Institutions*. Princeton, New Jersey: D. Van Nostrand Company, Inc., 1967 (\$2.95).

Normative data are being collected. So far, inter-judge reliabilities have ranged from .92 to .99.

The test is contained in the book mentioned above.

DONALD G. BARKER, Texas A and M University, College Station 77843

Scale of Attitudes Toward School Guidance

Forms A and B each contain 20 attitude statements arranged in an order of scale values that simplifies scoring but is not readily transparent to the respondent. The arrangement from most favorable to most unfavorable is of ascending scale value from item 11 through item 20 and then from item 1 through item 10.

Each respondent indicates his attitudes by placing a check mark by the statements with which he agrees. A space is provided for free-response remarks. Appropriate age level was not specified.

A subject's score is the median of the scale values of the attitude statements that he checks. If an odd number of items is checked, the score is simply the scale value of the middle item; when an even number of items is checked, the score is the mean of the two middle items. Most subjects check from three to seven items.

The Pearson product-moment coefficient of correlation of the scores on Form A and Form B was 0.709.

The instrument is reproduced in *Personnel and Guidance Journal*, June 1966.

DANIEL W. BEHRING, 3139-1 University Boulevard West, Kensington, Maryland 20795

Activities Preference Achievement Scale (APAS)

This scale has been used to aid in predicting high and low achievement among college freshmen at Ohio University and at Ripon College. It is a self-administered 60-item scale in which the subject registers "like," "indifferent," or "dislike" for each of 60 activities.

The following are examples of activities presented in the scale:

"Doing something that might provoke criticism."

"Planning my reading and outlining a reading program for myself."

"Wishing I could do something to someone to spoil his luck."

Two overlay keys, one for high- and one for low-achievers, are available. For each item all three possible responses are weighted in a range from zero, which is maximal discrimination of the low achievers, to 15, which is maximal discrimination of the high achievers.

Test-retest reliability for 128 freshmen at Ripon College was .73. Predictive validity, determined by correlation with grade point average, was .35 at Ripon College and .37 at Ohio University. A regression equation using APAS, high school rank, and CEEB-Verbal was developed for students at Ripon.

The instrument is available from the author for research purposes.

JOE M. BLACKBOURN, 1003 Vest Drive, Warrensburg, Missouri

Junior High School Articulation Scale

This instrument attempts to measure junior high school articulation problems as the students perceive them. Five subscales of the instrument are curriculum, interpersonal relationships, orientation, security, and personal counseling. Developed for a doctoral thesis, it was used to compare problem perception by sex, age, socioeconomic status, and schools.

The scale, requiring Likert-type responses, is virtually self-administering. Responses are scored one to five and the total score is the sum of the weighted responses. Means and standard deviations by sex, age, socioeconomic status, and schools are available for 1,671 students in seven junior high schools.

Content validity was established by a panel of five judges. No item was used unless it was selected unanimously by the judges. Factor analysis is being continued in an effort to develop more internally consistent scales.

The instrument is available from: University Microfilms, Inc., Ann Arbor, Michigan; also: Library, East Texas State University, Commerce.

OLIVER H. BROWN, University of Texas, Austin

Self-Report Inventory—Form R-3

This inventory consists of 48 statements which the subject rates on a five-point scale from "like me" to "unlike me." The following are examples of statements in the inventory:

"The way I get along with my friends is extremely important to me."

"I resist getting down to work and often have to drive myself to get it done."

"The sheer joy of being alive has often been a compelling force in my life."

For further information, contact the author.

PAUL B. CAMPBELL, Bureau of Quality Education Assessment, Room 570, Education Building, Box 911, Harrisburg, Pennsylvania 17126

Reading Attitude Inventory

This scale was devised to assess the effect of the reading techniques upon junior high school students' attitudes toward reading tasks. An elementary form is being developed.

Because it has been used only with remedial reading students in Livonia, Michigan, and with Title I evaluations, norms as such have not been established. There are, however, an item analysis and a reliability study.

Scoring was explained as five points for "the most positive response" and one for "the least positive." Summation of item scores provides an overall scale value.

Additional information may be obtained from: Dr. June Slobodian, 15125 Farmington Road, Livonia, Michigan 48154.

RICHARD E. CARNEY, Department of Psychology, California Western University, San Diego 92106

Achievement-Orientation Scale

This is a questionnaire designed to measure the degree of achievement motivation in normal human subjects. Derived by factor analysis from the California Psychological Inventory, it has been used with thousands of college students in experimental arousal studies and in correlations with such variables as sex, religion, social class, academic achievement, and other measures of motivation.

Standardized written instructions are given for true-false objective scoring which may be performed by hand or by machine.

Achievement-Orientation (AO) indicates the degree to which a person describes himself as dominant, independent, and achievement motivated. Groups traditionally more identified with intellectual endeavors are highest in AO and are more apt to control their perceptual environment. This scale, along with the other major CPI factor, social-orientation (SO), taps the central dimension of normal behavior motivation. The Eysenck Extroversion and Neuroticism Scales measure similar dimensions. See CPI references for normative data.

The scale is available from Consulting Psychologists' Press, Palo Alto, California.

COLLIER COUNTY BOARD OF PUBLIC INSTRUCTION, Collier County, Naples, Florida

"How I Feel" Attitude Inventory Test

This instrument assesses primary students' attitudes toward school and reading. It can be modified to measure attitudes toward many other things by simply changing the stimulus statements.

The Reading Inventory consists of 12 statements which are read to the students by the teacher. The following are examples of the statements:

"I feel this way when it is time for my reading lesson."

"I feel this way when my teacher chooses me to read aloud to the class."

"I feel this way when I meet new words while I am reading."

In response to each statement, the student circles the one of a set of six faces which portrays his feelings. In each set of six faces there are expressions of happiness, sleepiness, fear, anger, unhappiness, and indifference.

For further information contact Miss Catherine Archibald, Collier County Board of Public Instruction, Naples, Florida.

CLAUDE D. CUNNINGHAM, 2800 Steiner Street, San Francisco, California 94123

Test Attitude Scale

Using a Thurstone scale technique, this instrument attempts to measure the degree of test resistance among university undergraduates and high school seniors planning to attend college. It has been used by the author in only one study.

The 50 items of the scale are statements pertaining to feelings associated with test-takings. The following are sample items:

"It is good to have tests to give us information about people."

"I feel upset when I cannot answer a test question."

"I believe that students who score high on an intelligence test are favored by teachers."

Students respond by indicating agreement or disagreement with each statement. Scores are derived from a count of items which indicate test resistance.

Normative data have not yet been developed. Content validity of items was determined by a panel of judges composed of counselors enrolled in the NDEA Institute at Indiana University. An item analysis was performed with two groups of students to determine the most discriminating items.

The instrument may be used without charge for research purposes. It is available from the author.

WILLIAM W. FARQUHAR, 439 Erickson Hall, Michigan State University, East Lansing 48823

M-Scales

This is a group-administered, paper and pencil, self-rating scale which attempts to assess a student's attitudes toward academic tasks, reflected self concept, and selected personality traits. It is a research instrument and has been used to identify descriptive differences among Negro, Indian, parochial, and Jewish students. It has also been used to identify students with low motivation. It has been translated into Hebrew and is being translated into Spanish.

Normative data have been gathered on most of the samples listed above. The reliability for the subscale and total scales ranges from .68 to .92 for males and from .60 to .93 for females. For a sample of 254 males and 261 females, the correlations with grades were .56 and .40 respectively. Cross-validation estimates were .49 and .48 for males and females. The subscale correlations ranged from .27 to .42 for females and from .32 to .51 for males.

Scores involve a weight of 0-1 for each item, with a high score indicating high academic motivation. Seven factors have been identified to profile the components of academic motivation.

The instrument is available from the author.

JOHN W. FRENCH, New College, Sarasota, Florida 33578

Questionnaire on Motivation in College

Combining open-ended and structured responses, this instrument was used to find motivational types through inverse factor analysis. It has been used only in motivation research with college students; consequently there are no normative data.

3

The author asserts that this is *not* intended to be an operational instrument.

Items which elicit a structured response are in the form of phrases which the subject rates on a five-point scale according to importance. The following are illustrative of the 13 phrases:

To receive good grades (complete record, honors, favorable letters home, etc.)

To learn what is being presented

To prepare to do something that will improve the world.

Four open-ended questions follow the rating scale. Responses to these questions are analyzed by inverse factor analysis.

JACK R. FRYMIER, College of Education, The Ohio State University, Columbus 43210

Junior Index of Motivation—JIM Scale

The JIM Scale student questionnaire is an instrument for assessing students' motivation toward school. It has been used with over 15,000 students in several studies, some of which are described in: *The Nature of Educational Method*. Columbus, Ohio: Charles E. Merrill Books, Inc., 1965.

Eighty Likert-type items comprise the scale, but only 50 of them are scored. Students respond to each statement with +2 for strong support, agreement; +1 for slight support, agreement; -1 for slight opposition, disagreement; or -2 for strong opposition, disagreement. A student's scores for the 50 items are summed algebraically. Then the sign of the sum is reversed and it is added to 100 algebraically. Higher scores indicate higher motivational levels; low scores indicate low motivation levels.

Normative data are available for a stratified national sample of 3179 seventh through twelfth graders.

This instrument is available upon order (\$3.75 per 100 copies) from the publication office at: The Ohio State University, 242 West 18th Avenue, Columbus, Ohio 43210.

E. VAUGHN GULO, Northeastern University, Boston, Massachusetts

Attitudes Toward Professors

This is a modified version of a semantic differential which has been used by the author in several studies involving college students. It attempts to measure the multidimensional pattern of attitudes of students toward their ideal professors and their actual professors.

No statistical data are available; however, factor analysis has provided some information regarding validity.

Detailed descriptions of the technique are available in a dittoed article by Gulo, and the technique may be duplicated.

NASON HALL, Department of Sociology, The Ohio State University, Columbus 43210

The Junior High Boy

This self-rating checklist was used by the Youth Development Project in a study of delinquency proneness of boys 12 to 14 years of age. It consists of 33 direct questions about the individual's actions in relation to other people, to authority, and to private property.

Sample items:

"How often have you tried to stop a fight?"

"How often have you taken little things (worth less than \$2.00) that did not belong to you?"

"How often have you been really nice to one of your teachers?"

A student answers a question by checking the alternative which indicates how often since the beginning of the school year he has done what the question specifies. The number of alternatives varies from four, which is most frequently used, to seven, which is used for only three items.

For further information contact Nason Hall.

NASON HALL, Department of Sociology, The Ohio State University, Columbus 43210

Scales for School and Law Attitudes

This instrument of 135 items yields 16 scales such as "capacity to learn," "value of education," "policemen—relationships with kids," and "laws-legitimacy," to name a few. Students respond to statements by marking categories of strongly agree, agree, undecided, disagree, or strongly disagree. Responses are ranked from one to five or from five to one, depending upon the positive or negative function of the statement. The following are sample items from the scale:

"Everyone breaks the law from time to time."

"Don't let anybody your size get by with anything."

The instrument has been used only with boys of ages 12 to 14 from working-class families, but the author stated that it is appropriate for use with pupils in grades four through twelve.

For further information contact the author.

NEIL B. HOLLIMAN, Midwestern University, Psychology Department, Wichita Falls, Texas

Paired Comparison Technique

This is a psycho-physical scaling method adapted to establish reliable scale values for incentives based upon seven- and eleven-year-old children's judged preferences. Incentives used were bubble gum, candy, peanut, penny, balloon, marble, peg, bolt, tack, light, buzzer, and a word (yes). The incentives were glued on plain 4" x 6" index cards, except for the peg, bolt, and tack, which were presented on a circular block of wood. The light and buzzer were contained in a specially constructed apparatus. The incentives were presented in pairs, and the subject was requested to indicate which of the two he preferred. Combination and order of presentation were arranged and counterbalanced according to optimal methods worked out in previous psycho-physical studies using paired comparisons.

Reliability data for scales for seven- and eleven-year-old males and females are available. Validity studies relating scale values of incentives to ability to function as reinforcers for learning tasks show a positive general relationship.

The Guilford shortcut to the composite standard method was employed for scoring. The authors' research indicates that reliable scale values can be obtained by this technique from any representative group of age six or above.

A more exact description of the administration procedure will be furnished by the author upon request.

JOHN E. JORDAN, College of Education, Michigan State University, East Lansing

Attitudes Toward Education

This instrument, which has been used in several doctoral studies at MSU, is appropriate for use with adults. Form, administrative procedures, scoring procedures, normative data, and statistical data were not described. For details see: *Attitudes Toward Education and Physically Disabled Persons in Eleven Nations*, by John E. Jordan. Latin American Studies Center, Michigan State University, 1968. 312 pp.

LAWRENCE F. LOWERY, 4651 Tolman Hall, University of California, Berkeley 94720

The Projective Tests of Attitudes (PTOA)

Using three projective techniques—a word association test, a sentence completion test, and a thematic apperception test—this instrument

identifies the attitudes of children toward science and reading and may be used to uncover attitudes toward various other aspects of the curriculum. Some training is necessary for the use of this individual interview technique.

The word association test consists of 10 common nouns (e.g., house, dog, car), three of which pertain to science (science, experiment, and scientist). Responses are scored by judges.

The Lawrence Lowery Apperception Test provides drawings of neutral situations and explanatory statements with related questions which the respondent must answer. The drawings show a child in a situation pertaining to each of the study's basic themes (science, process, scientist). Responses are scored by judges.

In the sentence completion test, respondents are asked to complete nine statements such as the following:

"The field of science is _____."

"Most people like science whenever it _____."

Responses to all three of the tests are scored as positive, neutral, or negative by the judges. No normative data are available, but statistical data are available in an article in *School Science and Mathematics*.

Because this is primarily an experimental research instrument, information can be obtained only from the author.

BOYD R. McCANDLESS, Department of Psychology, Emory University, Atlanta, Georgia

Intensity of Involvement Scale (Observation)

This observational method has been used with four- and five-year-olds in teacher-structured situations and may be equally useful in free play situations. It entails observations five seconds in length which are then categorized according to six subjectively identified degrees of involvement. The six categories are described in behavioral terms to guide the observer, who records a number for each observation period. Briefly, the six categories are "unoccupied," "onlooking," "minimal-minimal," "minimal," "attention moderate," and "complete." A scoring sheet was designed to facilitate recording.

The scale was designed for research purposes only. The authors reported good luck establishing reliability in between one to three hours of dyadic or triadic reliability training (percentages of agreement up to 96 with correspondingly high relationship quotients) and the belief that a reliable sample of a child's behavior may be a useful predictor of his later adjustment to school.

WALLACE H. MAW, University of Delaware, Newark, and ETHEL W. MAW, Bryn Mawr College, Bryn Mawr, Pennsylvania

The You Test

"The You Test" is composed of three subtests, "Which Do You Think Are Foolish Sayings?"; "What Do You Know?"; and "About Myself." In the first subtest, which consists of 22 items, students are directed to put an X before statements which "have parts in them that make them foolish" and a C before those which are "all right." Examples of statements are:

"The soldiers were outnumbered so they gave up without a fight."

"They picked up the melted ice cubes and dropped them into the pail."

"If you can't read type of this size, you need glasses."

In the second subtest the students are to respond to 17 factual multiple-choice items on a number of subjects. An example of these items is:

The wife of John Adams was a person of great ability. Her name was

_____ Abigail

_____ Audrey

_____ Alice

_____ Aletha

The third subtest is a self-report checklist of 41 statements. Each statement requires a response of never, sometimes, often, or always. Examples of statements are:

"I like to explore strange places."

"I keep my hands clean."

"I find that things puzzle me."

For further information contact the authors.

PAUL A. NELSON, 410 Glenn Drive, Urbana, Illinois 61801

Content Attitude Test

This Likert-type scale investigates teacher attitudes toward the developmental potential—both academic and social—of various elementary content areas. It is appropriate for use with preservice and experienced teachers. The author hopes that attitudes toward content areas can be improved if they are once identified.

The scale consists of 76 items to which the subject responds on seven-point scales. The following are items taken from the scale:

1. To what extent does the teaching-learning situation in social studies provide opportunity to develop creative thinking?

12. To what extent does the teaching-learning situation in science provide opportunity to use a demonstration-discussion approach to teaching?

48. To what extent does the teaching-learning situation in arithmetic lend itself to the homogeneous grouping of students by classes?

75. To what extent does the teaching-learning situation in language arts provide opportunity for the development of leadership-membership abilities?

Contact the author for further information.

WILLIAM FRANK ROWE, 487 Steeple Chase Lane, Somerville, New Jersey 08876

School Attitude Q-Sort

The purpose of this technique is to determine the subject's attitudes toward schooling, authority-discipline, and schoolwork. Although the technique has been used only with junior and senior high school students, it is also appropriate for elementary students.

The Q-sort contains 60 items. A 20-item school attitude questionnaire was used in conjunction with the Q-sort.

For more information contact the author.

GEORGE E. SCHLESSER, 28 Payne Street, Hamilton, New York 13346

Personal Values Inventory

This instrument is designed to measure academic motivation among college freshmen. It has been used in several dropout, follow-up, and pattern studies since 1961. Norms for eight subtests are available for men and women in two- and four-year colleges. Scores correlate .4 to .5 with first semester grade point average, and reliability is satisfactory.

A self-administering test, the "Personal Values Inventory" requires about 45 minutes. Scoring by computer is available at Colgate University. Colleges must cooperate in the research to take advantage of the service.

The instrument costs 20¢ per student, including scoring and reporting. Write to the author for further information and copies of the inventory.

GLEN ROBBINS THOMPSON, 9445 Gross Point Road, Skokie, Illinois 60076

Preschool Academic Sentiment Scale (PASS)

The purpose of the instrument is to assess the attitudes of pre-school and young school children toward learning and school. PASS has been used experimentally to evaluate the effect of Title I programs on the academic attitudes of young school children.

Normative data are presently being gathered for urban and suburban children ages 4-6 and 6-7. Statistical data are not yet available.

PASS may be administered by untrained personnel to groups of children who respond nonverbally to stimulus statements read by the tester. Procedures are detailed in the manual. Special instructions are given for the use of PASS with educationally disadvantaged children. Clerical personnel may score the test.

Final editions should be available in 1968 from the following: Priority Innovations, Inc., P. O. Box 792, Skokie, Illinois 60076. Experimental edition specimen set: \$2.50.

FRANK H. WOOD, 103 Pattee Hall, University of Minnesota, Minneapolis 55455

Experimental Procedure for Measuring Reading Achievement Motivation in Children

Based on Crandall's theory of children's achievement motivation, this technique was designed to measure (a) reading achievement expectancy and (b) attainment value placed on successful reading achievement. It was developed to obtain more than a dichotomous (yes-no) response from children of ages four to seven replying to questions about attitudes and values, and it requires no verbal responses. It has been used with culturally disadvantaged white, Negro, and Indian children from inner-city, low-income neighborhoods.

The technique requires the child to position four cards picturing a child on each (two boys and two girls) in answer to questions posed by the examiner. The examiner proceeds in the following manner: Using the questions specified, he asks the child which one of the children in the pictures wants to do something, e.g., "learn to read with the group." The experimenter places this picture on the top step and proceeds to ask the child which of the children in the pictures that are left wants to "learn to read with this group" the most. This picture is placed on the second step, and the other two pictures are treated in a similar manner asking finally which wants to do the activity least. Then the experimenter asks the child to indicate which of the children pictured and positioned is most like him (the child).

The order of the figures on the steps from top to bottom is noted by the examiner, e.g., L-B-G-G and the position identified by the subject as most like him is circled. Scoring is, from top to bottom step, 4-3-2-1. The figure cards are taken down and rearranged for each question. The following are some other questions from the instrument:

(2A) Which one of the children wants most to answer questions about what they read in their books?

(2B) Which one of the children is best at answering questions about what they read in their books?

Normative data for this technique are limited to the samples of disadvantaged children mentioned above. Statistical data are currently being compiled from first-grade usage. Reliability has been found to be low but adequate. Using reading achievement test performance and teacher rating of reading achievement efforts as criteria, validity has not been satisfactorily established. The author feels, however, that the procedure itself shows promise and hopes to replicate the study with a sample large enough to permit differentiation into ability groups as well as sex and race groups.

If contacted, the author will provide further information.

LAWRENCE S. WRIGHTSMAN, George Peabody College, Box 512, Nashville, Tennessee 37203

School Morale Scale

For use with students in grades four through twelve, this instrument assesses seven different areas of school morale—School Building, Instruction and Instructional Materials, Teacher-Student Relationships, Community Support and Parental Involvement, Relationships with Other Students, Administration and Regulations, and General Feelings About School. It has been used to evaluate ESEA Title III projects, and normative data are available on several hundred students in each grade.

The scale is composed of 84 items for which the student marks either "agree" or "disagree." It can be administered to groups, but with younger children (ages 10 to 12) it is best to read each item to the group.

Hand-scoring sheets are available. Scores range from 0-12 on each of the seven subscales.

The School Morale Scale is available from the author.

JOSEPH C. ZINKER, 19407 Wickfield Avenue, Cleveland, Ohio 44122

Q-Sort for the Hierarchy of Needs

The purpose of this Q-sort is to evaluate levels of motivation (in terms of Maslow's need hierarchy) in physically ill hospitalized patients. It has been used by several investigators, and is appropriate for disabled and older individuals as well as for physically ill persons.

Normative data are limited.

For information on statistical data, administration procedures, scoring, and analysis, see: J. C. Zinker. *Rosa Lee: Motivation in the Crisis of Dying*. Painesville, Ohio: Lake Erie College Press.

The instrument is available from: Lake Erie College Press, Lake Erie College, Painesville, Ohio. Price: \$2.50.

Personality

S. MARY AMATORA, Tests and Services Associates, 120 Detzel Place, Cincinnati, Ohio 45219

Personality Rating Scale

This scale may be used with all school age subjects (K-12) to appraise 22 areas of personality chosen from other available scales and from traits suggested by colleagues. The wording of items is at the third-grade level so that most subjects may complete the scale unaided; however, below the third grade it is necessary for the administrator to ask the subjects questions and record their responses.

Each student rates several of his classmates (the particular number is set by the teacher after considering the time available) and himself. Then each student is given all ratings made of him (in lower grades adults must score the forms); he averages the ratings made by boys and those made by girls; and he constructs a personality profile for himself.

Reliability coefficients are reported for each of the 22 traits. Ratings by girls are slightly more reliable than by boys, but there is a wide range of reliability (.40-.86). Validity is claimed on the basis of agreement between self-evaluations and opinions of classmates.

In the third grade or above, the scale can be group-administered in 30 to 40 minutes with each child rating 10 to 15 others.

The scale consists of one item per trait. Items are in the form of questions, and responses are based on 10-point continua.

Responses are scored on grid sheets to facilitate distribution of ratings. The following are examples of items:

(pep)	gets tired easily				average		peppy		
1. Is he "peppy" and full of life?	1	2	3	4	5	6	7	8	9 10
(intelligence)	very dull				average		very bright		
2. How bright or intelligent is he?	1	2	3	4	5	6	7	8	9 10
(courtesy)	very impolite				average		always polite		
3. How polite and well-mannered is he?	1	2	3	4	5	6	7	8	9 10

The Personality Rating Scale may be ordered from the above address. Prices: \$3.50 per pkg. (35 tests, 35 pupil rating sheets, three class record sheets, one key, and one manual). Specimen set—60¢.

KATHERINE BEMIS, Department of Education, University of New Mexico, Albuquerque

Teacher Observation Personality Schedule (TOPS)

The purpose of this instrument is to measure classroom behavior which seems related to the Edwards' Personality Preference Schedule needs of achievement, abasement, affiliation, orderliness, change, dominance, and heterosexuality. It was standardized in 1965 on a sample of eight teachers, each observed two times, and it was used by Cooper and Bemis (1967) to observe each of 60 fourth-grade teachers nine times. Data from both of these studies are available. Inter-rater reliabilities have been satisfactory.

Observations using TOPS require 22½ minutes. The trait heterosexuality is observed for 7½ minutes, and the other traits are observed for a total of 15 minutes, alternating each five minutes, using the sign system given on page two of the schedule.

For each observational category, a total score is obtained by summing the total number of times a particular behavior occurred. This procedure yields 60 scores per teacher. The scores lend themselves to analyses of teacher behavior such as factor analysis and canonical correlation.

For further explanation of this technique, contact: Dr. James G. Cooper, College of Education, University of New Mexico, Albuquerque, New Mexico.

DONALD A. BLOCH, M.D., 149 East 78th Street, New York, New York

Deviant Behavior Inventory for Children

The purpose of this instrument is to review the total range of deviant behavior and assess its presence or absence in accordance with a technique for recording false positives, false negatives, and true deviant behavior. The instrument was developed primarily for use with latency age children.

Statistical information can be obtained through Dr. Eva Rosenfield, Jewish Board of Guardians, 120 West 57th Street, New York, New York.

EDGAR F. BORGATTA, University of Wisconsin, Madison

Behavioral Self-Rating Form (BSR)

The BSR provides a simple, direct personality measure of college students, but its use is recommended only for situations in which a *very*

short test of personality can be useful. It can be used effectively to provide additional scores in more extensive personality testing to permit examination of content validity.

Norms are available in the form of decile distributions for college samples. Validity is indicated through prediction of parallel peer assessments in a multitrait-multimethod matrix approach.

The scale consists of 20 statements which the subject rates along a 10-point scale from "Definitely does not describe me" to "Definitely describes me well." The following is a sample taken from the scale.

	Definitely does not describe me					Definitely describes me well				
Generally, I . . .										
am very active	0	1	2	3	4	5	6	7	8	9
am friendly	0	1	2	3	4	5	6	7	8	9
am intelligent	0	1	2	3	4	5	6	7	8	9
am very tense	0	1	2	3	4	5	6	7	8	9

The data from samples support the view that this type of rating has internal consistency within five differentiated content areas—assertiveness, likeability, emotionality, intelligence, and responsibility. The first two areas are simple sums of three items each, while scores for the others are simple sums of four items each.

The instrument is available from the author.

EDGAR F. BORGATTA, University of Wisconsin, Madison

A Short Test of Personality: The S-ident Form

The S-ident Form is a 31-item, self-report inventory of personality used with college students and high school juniors. It consists of six subtests which are based on factor and cluster analyses of 114 personality items. The six personality scores—leadership, impulsiveness, intellectual interest, aloofness, self-depreciation, lack of tension, and emotional instability—appear to correlate substantially with at least one factorially based peer-ranking score.

To each of the 31 items, the subject may respond with one of four alternatives—definitely agree, probably agree, probably disagree, definitely disagree.

Sample items:

I make friends easily.

I usually act on the spur of the moment.

I often feel that I have more problems than other people.

Contact the author for further information.

LEONARD P. CAMPOS, OH Close, School for Boys, Box 6500, Stockton, California

Story Completion Technique—Children's Form

This technique measures the ability to delay need gratification in children aged 8 to 13. The child is asked to complete (in oral or written form) story situations which present opportunities for different degrees of gratification latency on five needs—acquisition, aggression, nurturance, achievement, and affiliation. A scale of from one to seven points is employed to represent gratification of need from most immediate to most delayed.

Since it was used only in a doctoral research project, there are no normative data. Interscorer reliability was very high (.90 and above). Construct validity determined by correlation with Rorschach was statistically significant (.05).

Further details of the technique and rating procedures are available from the author.

RAYMOND B. CATTELL, Department of Psychology, University of Illinois, Urbana

High School Personality Questionnaire (HSPQ)

The HSPQ is a questionnaire designed to aid teachers, guidance specialists, researchers, and clinicians by giving a standardized objective assessment of an individual's general personality. It measures 14 personality dimensions that were selected because they come near to covering the total personality. (See 16PF.) The test is applicable to those from 12 to 14 years of age. Scoring is rapidly accomplished through a stencil key. The test is available in forms A and B.

Satisfactory reliability coefficients for the 14 dimensions were reported. Construct validity was claimed on the basis of multiple correlations between item and factor of the 14 dimensions. The test can be administered individually or to groups, and no time limit is employed, although all but the slowest readers can complete the test in 40 to 50 minutes. Scores are reported in stens. Further information is available from the Institute of Personality and Ability Testing, 1602-04 Coronado Drive, Champaign, Illinois.

RAYMOND B. CATTELL and H. W. EBER, IPAT, 1602-04 Coronado Drive, Champaign, Illinois

The Sixteen Personality Factor Test (16PF) Forms A, B, C

As a factor analytically developed personality questionnaire, the 16PF is designed to measure the major dimensions of personality com-

prehensively. Originally developed in 1950, the 16PF has been repeatedly revised. The test is applicable to those included in the age range from 16 years to late maturity. Simple administration procedures, expedient scoring (stencil key or machine), and its wide range of applicability make the test extremely usable. Normative data are presently based on six or seven thousand cases and are continually being updated.

Split-half reliabilities range between $+.71$ to $+.93$, the average being about $+.83$ to $+.84$. Internal construct validity averages approximately $+.88$ and ranges from $+.73$ to $+.96$. The test scores are very useful when multiple correlations are to be utilized because of their comprehensive nature. Comprehensive clinical and occupational criterion groups are available for comparison.

The test gives scores on 16 primary dimensions of personality as well as four composite "second-order" scores from combinations of the primary factors. Examples of the primary dimensions are: Reserved vs. Outgoing, Humble vs. Assertive, Shy vs. Venturesome, Practical vs. Imaginative, Placid vs. Apprehensive, Group Dependent vs. Self Sufficient. The composite second order scores are reported on: Anxiety, Extroversion vs. Introversion, Tough Poise vs. Emotionality, and Independence vs. Dependence.

RAYMOND B. CATTELL and IVAN H. SCHEIER, IPAT, 1602-04 Coronado Drive, Champaign, Illinois

IPAT Anxiety Scale Questionnaire

This scale is a brief, valid, and nonstressful questionnaire designed to measure anxiety level in adults and young adults down to age 14 or 15. The test adapts itself for either group or individual administration and can be taken without the presence of an administrator (thereby reducing stress). The scale is composed of 40 items that are distributed over five anxiety measuring factors (Defective Integration, Ego Weakness, Suspiciousness, Guilt Proneness, and Frustrative Tension). The test can be utilized as a screening method as well as in clinical settings.

Construct validity is estimated at $.85$ to $.90$ for the total scale. Concrete validity seems substantiated by consensus of clinicians ($+.30$ to $+.40$). Dependability reliability ranges from $.87$ to $.93$ for test-retest, and homogeneity coefficients range from $.80$ to $.91$. The test utilizes sten scores for normative comparisons. Experimental covert and overt subscores are being developed.

RICHARD W. COAN, University of Arizona, Tucson, and RAYMOND B. CATTELL, University of Illinois, Urbana

Early School Personality Questionnaire (ESPQ)

Here is a personality instrument designed for the six- through eight-year age group. It offers measures on 13 personality dimensions (see 16PF) with a minimum of testing time. The administrator reads questions (80 items) aloud and the subjects record their responses on a nonverbal answer sheet. The test may be given individually or in groups of 20 to 30, and little advance preparation is necessary. Scoring is performed by a stencil that is provided with the test.

Scores are reported in stens and appropriate norms are included with the test. Information on validity and reliability can be obtained from IPAT, 1602-04 Coronado Drive, Champaign, Illinois.

DWIGHT G. DEAN, Denison University, Granville, Ohio 43023

Alienation Scale

The Alienation Scale was designed to yield separate scores for each of three dimensions of the alienation syndrome—sense of powerlessness, normlessness, and social isolation. It is appropriate for use with high school students and adults. It has been used in studies of alcoholism, crime, marriage, college dropouts, and others.

The scale consists of 24 items which employ a Likert scoring system. Typical of the nine items in the powerlessness scale were:

There is little or nothing I can do toward preventing a major "shooting" war.

We are just so many cogs in the machinery of life.

Typical of the six items in the normlessness scale were:

The end often justifies the means.

I often wonder what the meaning of life is.

Typical of the nine items of the social isolation subscale were:

Sometimes I feel all alone in the world.

One can always find friends if he shows himself friendly.

Normative data (means and standard deviations) are available for a variety of samples. Reliabilities of subscales were reported as satisfactory as was that of the total scale. Consensual, independent judgments of seven experts were cited as evidence of validity.

The scale is available through the author.

DWIGHT G. DEAN, Denison University, Granville, Ohio 43023

Emotional Maturity Scale

This scale measures 14 different components of emotional maturity including stress, anger, authority, intellectual maturity, and social poise.

In the past the instrument has been used to correlate emotional maturity with marital adjustment, to predict success in college, and to correlate emotional maturity with support of school bond issues. It was used with high school, college, and other samples, and some normative data are available in article reprints.

Reliability (split-half) was reported as .75. Validity was asserted on the basis of (a) items derived from Eibert, who checked with 14 psychiatrists; (b) comparisons between freshmen men and freshmen women and between freshmen women and senior women which yielded differences in the predicted direction; and (c) ratings of girls by housemothers.

Likert-type scoring was used with statements for which the respondent was to rate himself. The following are examples of the 183 statements in the scale:

- 2. Remains cheerful even when things aren't going his way.
- 8. Seems emotionally secure, seldom exhibits anxiety.
- 22. May provoke others unnecessarily.

The instrument is available through the author.

H. J. EYSENCK, Department of Psychology, Decresigny Park, London, England

Junior Eysenck Personality Inventory

This scale is designed to measure the two major personality variables of neuroticism, or emotionality, and extroversion/introversion in children. It was developed from the Maudsley Personality Inventory and the Eysenck Personality Inventory for adults. The test consists of 60 yes and no questions, and has three scales for which scores are reported (N-scale or Neuroticism, E-scale or Extroversion, and L-scale or Lie scale). The Inventory is appropriate for children aged 7 to 14.

Reliability measures are reported for test-retest (+.70 to +.80) and for split-half (+.75 to +.85). The data on the validity of the scale are incomplete at present. Norms are available and scoring is accomplished through either a scoring key or computer.

Examples of items:

	yes	no
Do you like plenty of excitement going on around you?	0	0
Do you often need kind friends to cheer you up?	0	0
Do you nearly always have a quick answer when people talk to you?	0	0

The *Junior Eysenck Personality Inventory* is available from: University of London Press, St. Paul's House, Warwick Lane, London, E.C. 4.

H. B. GIBSON, Institute of Criminology, University of Cambridge, England

The Gibson Spiral Maze

The Maze (1965) is a simple, individually-administered psychomotor test which enables one to measure speed and accuracy simultaneously. Originally designed for research work with school children, it has since been used for a wide variety of purposes with all age levels. Some of its uses have been to assess psychomotor dysfunction in research into delinquent maladjustment, to give a *personality measure* related to extroversion and neuroticism, and to make an index of improvement in the psychiatric treatment of depression.

Printed on a 10-inch square card, the maze has to be negotiated by pencil as quickly as possible, avoiding round obstacles drawn inside the maze. Error points are given for touching the obstacles or the sides of the maze.

Norms for schoolboys in the age range of eight to ten years are included in the manual. Additional information on norms derived from various other uses of the test will be published at a later date.

Cited in the manual as evidence of the validity of the test are the following:

1. Different sorts of people characteristically give different types of performance, e.g., "good" and "naughty" boys.
2. Certain social characteristics of primary schoolboys are significantly associated with poor performance on the maze.
3. The adjusted error score correlates (.33) significantly with the Porteus Q Score which is minimally loaded with intellectual ability.

The simplicity of administration of the maze calls for utmost rigour in standard administration. Only qualified psychologists should attempt its administration and interpretation of its results. Results should not be expected to be entirely revealing in themselves, but should be used in conjunction with various other instruments and techniques.

The Gibson Spiral Maze is available from: University of London Press, Ltd., Saint Paul's House, Warwick Lane, London E.C. 4. Manual—four shillings each; test cards—four pence per copy.

H. B. GIBSON, Institute of Criminology, University of Cambridge, England, and W. D. FURNEAUX, Department of Psychology, University of London, England

The New Junior Maudsley Inventory

The New Junior Maudsley, formerly the Junior Maudsley Personality Inventory, is concerned with the assessment of personalities.

It has been used in many educational and clinical studies. (See *British Journal of Educational Psychology* since 1961.) The test contains 64 questions that require either a "same" or "different" response. It yields scores on three scales: Extroversion (E), Neuroticism (N) and Lying (L). The scale is appropriate for those between the ages of 9 and 16 years.

Reliability measures for the E and N scales range from $+.74$ to $+.94$ (Guttman, split-half, and test-retest). Validity is reported by mean item/scale correlations and these range from $.42$ to $.52$. Correlations between the N and E scales are about $+.10$. The norms were derived directly from the JMPI except for the L scale. The Inventory can be administered individually or to groups and scoring is achieved through prepared templates or *ad hoc* analysis of data obtained from large groups.

Sample items:

- 3. I like to be in school plays.
- 13. I like to tell my friends all about things that happen to me.
- 23. I start the fun at a quiet party.
- 33. I like to work alone.
- 43. I often think people follow me at night.
- 53. I try to stop other children from using bad language.

Further information may be obtained through Educational and Industrial Testing Service, P.O. 7234, San Diego, California.

ALBIN R. GILBERT, 25 College Avenue, Buckhannon, West Virginia

Latency-weighted Personality Testing (Technique)

This technique is being developed to weight self-report responses to paper and pencil tests by the latency in making these responses. It has been used for testing of nonintellective traits of college students, of preministerial students, and of mental patients. It can be used for any group amenable to paper and pencil inventories and has been used on college and high school students, mental patients, and vocational applicants. No normative data are yet available.

Split-half reliability is reported at $+.92$. Validity (concurrent) studies are under way but results are as yet unavailable. It was found, however, that latency-weighted scores discriminate more sharply than unweighted paper and pencil scores. Administration involves a tape upon which the paper and pencil items are recorded. The latency is measured by a timing device that is triggered by the last word of each question and stopped by the response of the subject. Scoring is achieved by a profile of weighted and unweighted scores. Additional information may be obtained from Albin Gilbert.

PAUL KLINE, Institute of Education, University of Exeter, Exeter, Devon, England

Ai3

This paper and pencil test is designed as a measure of Freudian Anal Character. Although it is an experimental test being used primarily to investigate Freudian psychological hypotheses, the results, to date, show it to be a promising instrument. The test has been factor analyzed with the 16 PF, MMPI, and D.P.I., correlated with the Peabody Pictures, and used in the Coon Cultural Studies; however, no data or results are available at present. The test consists of 30 yes and no questions with a 10 minute maximum time limit. The Ai3 can be used with adults, 16 and over, if they are literate (a lower verbal level form is being developed).

Reliability centers around $+.67$ with a discrimination index (Ferguson's Delta) of $+.96$. Validity and normative data are incomplete but were to have been developed by the end of 1968. Scoring is performed with a scoring key supplied with the instrument.

Sample items:

Do you keep careful accounts of the money you spend?	Yes	No
When eating out do you wonder what the kitchens are like?	Yes	No
Do you insist on paying back even small trivial debts?	Yes	No

This instrument is available from the author.

PAUL G. LIBERTY, Southwestern Cooperative Educational Laboratory, 117 Richmond, N.E., Albuquerque, New Mexico 87112

SWCEL Student Questionnaire

This instrument is an interview technique for assessing "non-cognitive (personality, motivation) characteristics of first-grade pupils." It consists of six parts—(a) test/school anxiety; (b) sex role identification; (c) self-esteem; (d) acquiescence; (e) gratification delay; and (f) individual mastery. It has been used with 300 lower-middle class Anglo-American and Spanish-American children.

Normative data are being compiled, and statistical data are being analyzed. The instrument is considered satisfactory for exploratory research but is still in the process of being refined.

Responses to the items are either yes-no or very short answers. The interviewer records responses directly on the questionnaire form.

Sample items:

Do you like to take toys to school and show them to the children?
Do you think you will pass to the second grade?

Each circle stands for some person. Which one are you?

0 0 0 0 0

Would you rather have a penny today or wait until tomorrow for 5¢?

This instrument is available from the author.

RUSSELL E. MASON, 45 Alhambra Court, Portola Valley, California 94025

Cross-Cultural Functional Personality Analysis Inventory

This self-rating inventory was designed to encompass comprehensively and differentiate meaningfully the most significant aspects of human personality using group administration and rapid scoring in order to make cross-cultural comparisons. Although still in the process of development, the instrument has been used in studying personality dynamics and in a series of group sessions oriented toward self-understanding. It is appropriate for adolescents and adults, but normative data, which will initially include only college students, have not yet been published. Scoring procedures vary with each inventory scale (Development of Self-System—102 items, Responsible Persons' Attitudes—44 items, Self-System Identification—60 items, and Directive Control—45 items) and personality groupings of scales.

The instrument was to be made available at the end of 1968, for research purposes only.

For more details on the construction and underlying theory of the instrument, consult *Psychological Reports*, 1966, pp. 1179-1182; or contact the author.

ROBERT PLUTCHIK, Department of Psychiatry, College of Physicians and Surgeons, Columbia University, 722 West 168th Street, New York, New York 10032

Emotions Profile Index (E.P.I.)

The E.P.I. is a self-administering, forced-choice personality test which evaluates the relative importance of certain basic emotions in the life of an individual. The categories of analysis of the test are based on a theory of emotion which postulates that personality traits may be conceptualized as mixtures of two or more of eight basic emotions. The trait choices are scored in terms of the underlying emotions, thus producing a profile based upon eight emotion categories. It has been used with a variety of subjects including high school students, housewives, and drug addicts. It has also been used on a repetitive basis to chart the course of manic-depressive fluctuations and by teachers to evaluate emotional tendencies in elementary school children.

Short term test-retest reliability is over $+.90$ on all scales, and six-month test-retest reliability is between $+.6$ and $+.8$ on all scales. Validity is still being examined by "various methods."

This instrument is available from Dr. Robert Plutchik at the above address. Price: \$20.00 per 100 copies plus manual. (Sold at cost for research purposes only.)

RUTHERFORD B. PORTER, Indiana State University, Terre Haute,
and RAYMOND B. CATTELL, University of Illinois, Urbana

Children's Personality Questionnaire (CPQ)

This instrument is designed to yield a general assessment of personality development by measuring 14 personality dimensions thought by psychologists to approach the total personality. The test is applicable to those of ages 8 through 12. Scoring is accomplished with a stencil key, and appropriate sex and age norms are included with the test.

The test is administered without a time limit. It is available in Forms A and B, each form having two parts. Scores are reported in tens. Further information may be obtained from IPAT, 1602-04 Coronado Drive, Champaign, Illinois.

IVAN H. SCHEIER and RAYMOND B. CATTELL, IPAT, 1602-04 Coronado Drive, Champaign, Illinois

Neuroticism Scale Questionnaire (NSQ)

The NSQ is a test designed to measure the degree of neuroticism or neurotic trend in normal and abnormal adults and adolescents. It is a brief, standard, easily administered and scored part of the IPAT plan for providing measures for each factored personality dimension. The test provides measures on four components of the scale—Protected Emotional, Sensitivity, Depression, Submissiveness and Anxiety.

Split-half consistency coefficients of the four components of the scale range from $+.47$ to $.70$ but are generally in the order of $+.60$ to $+.70$. Construct validity is reported as ranging from $.69$ to $.84$ for the components, and concrete validity has proved to be substantial (one study reported that neurotics scored significantly higher than normals at $.0005$ level). The test also provides for profile analysis in relation to standard criterion profiles, both clinical and occupational. The test can be administered to groups or individuals.

GARY E. STOLLAH, Department of Psychology, Michigan State University, East Lansing

Problem List

This checklist of 237 child problems has been used in child psychopathology and psychotherapy, mainly in research studies and in pre- and post-therapy. Normative data are presently being accumulated. It is self-administrative and scoring is simply a tabulation of problems checked.

No statistical data on reliability and validity are yet available. Further questions should be addressed to the author.

Readiness

HARRY E. ANDERSON, University of Georgia, Athens 30601

Behavioral Maturity Scale

This rating scale measures academic, social, and emotional maturity. It consists of six items for each of the three factors. Teachers rate their students on the three factors. The score for each factor is the sum of the ratings (one to seven) on the items. The scale was used with Japanese and South Korean second graders as well as with four-, five-, and six-year-olds in the United States.

The author refers the interested reader to an article in a 1968 issue of *Educational and Psychological Measurements*.

TERRY DENNY, Coordinator of Field Research, Educational Products Information Exchange, New York

Reading Percepts Interview Schedule

This is an information-gathering technique designed to help assess children's perceptions of the reading act. The instrument can be used with children ranging in age from five to eight years. The schedule has been used in a longitudinal study of first and second graders but is definitely still in the research and development stage.

Small group pilot studies have indicated moderate to good inter-rater and intra-student reliability. Inter-rater agreement centers around the middle .80's and test-retest has yielded a .84. No validity measures or normative data have been compiled. The schedule is individually administered and the interviews range from 15 to 40 minutes in length. Interviews are scored by empirically derived categories.

The instrument is available from Dr. Samuel Weintraub, Department of Education, University of Chicago, who would consider sharing protocol only with others attempting to develop a technique with a similar purpose.

SAN FRANCISCO STATE COLLEGE, San Francisco, California

Levine-Elzey Preschool Social Competency Scale

The purpose of this scale is to measure the social competence of preschool children between the ages of two years six months and five years six months who do not have severe hearing, visual, motor, or emotional problems. Each child is rated according to how he actually performs at the time of rating, not according to how he might behave if conditions permitted.

Recorded on IBM cards, the first four items are for personal data (sex, chronological age, occupation of parents, and length of preschool experience), and the following 30 items comprise the rating scale. Each item in the scale contains four levels which are scaled from low competence (level A) to high competence (level D). The levels are cumulative, in that a child rated at the D level, for example, is presumed to be able to perform all preceding levels.

The following are sample items taken from the scale:

5. Identification

- A. Can state first name only.
- B. Can state full name.
- C. Can state full name and age as of last birthday.
- D. Can state name, age, and address.

15. Making Explanation to Other Children

When attempting to explain how to do something to another child (put things together, play a game, etc.)——

- A. He is unable to do so.
- B. He gives an incomplete explanation.
- C. He gives a complete but general explanation.
- D. He gives a complete explanation with specific details.

The scale, which is being supported by a grant from the U.S. Office of Education, has not been completely developed at this time.

Self Concept

CINCINNATI PUBLIC SCHOOLS, Division of Psychological Services
and Division of Program Development, Cincinnati, Ohio

What I Am Like

This is a self-rating scale based on Osgood's concept of the semantic differential. It is strictly experimental and should be used only for group comparisons, not for individual pupil diagnosis.

The instrument consists of three subtests, each containing 10 items. The first, *What I Look Like*, consists of adjectives characterizing physical attributes (short-tall, clean-dirty, awake-sleepy, etc.). The second, *What I Am*, attempts to measure self-image from a psychological point of view (happy-sad, somebody-nobody, bold-shy, etc.). The third, *What I Am Like When I Am with My Friends*, concerns social attributes (give-receive, agree-fight, follower-leader, etc.). Five-point bipolar scales are used in each subtest. The position of positive and negative poles was randomized to avoid a psychological set in rating items.

This instrument was used with 847 pupils in grades four through nine. It was viewed as having construct validity, but predictive validity has not been established. It should not be considered generally reliable for individual diagnosis.

STANLEY COOPERSMITH, Department of Psychology, University of California, Davis

Self-Esteem Inventory

This 58-item inventory is a method of studying self concept. In addition to a lie scale, the Self-Esteem Inventory has four subscales—self, social, home, and school.

Students respond to simple declarative sentences by checking "like me" or "unlike me" columns. The test is scored by totaling the "like me" and "unlike me" responses for each of the four scales and then adding these together. Norm tables by grade levels are available.

Sample items:

1. I spend a lot of time daydreaming.
15. Someone always has to tell me what to do.
30. It's pretty tough to be me.
45. If I have something to say, I usually say it.

Further information is probably available from the author.

CLIFFORD C. COURSON, Brevard Junior College, Cocoa, Florida 32922

Inference

This is a method for using inference as research data in the behavioral sciences. It is based on inference drawn from pairs of essays written by each of 64 senior high school students.

Students wrote two essays each; the first was entitled "A Teenager's Advice to the World," and the second was written in response to one of three selected pictures from the Thematic Apperception Test. Each essay was written during a 50-minute class period. Papers were collected and sent to a typist who transcribed the essays, omitting all identifying information.

A Perceptual Factors Rating Scale was devised to quantify the inferences made by raters about each subject for four perceptual variables identified by Combs as underlying the adequate personality. These factors were an essentially positive view of self, a feeling of wide identification with others, an openness to experience, and a summation of the foregoing three factors.

Inter-rater and intra-rater reliabilities were found statistically significant, and the author concluded that inference is a promising tool for gathering research data, especially on self concept.

Details of the procedures are published in *Educational and Psychological Measurement* 25: 1029-37, Winter 1965.

CHARLES VAN LOAN DEDRICK, 1420 Shady Acres Lane, Apopka, Florida

Perceptual Score Sheet

Dedrick submitted a perceptual score sheet to measure self concept on the basis of critical incidents and TAT interviews. It is presently being used to measure self concept of student and professional nurses and is considered relevant for all members of the helping professions (teachers, ministers, counselors, and nurses).

Although the Score Sheet is completed, statistical data are not yet available.

The Score Sheet and explanation of its four dimensions are mimeographed and may be obtained from the author.

DR. PRATIBHA DEO, Punjab University, Sector 14, Chandigarh, India: Department of Education

Personality Word List

The PWL was designed to measure self concept and specifically three aspects of self—perceived, ideal, and real. Finalized after three years of research, this instrument has been efficaciously used in many studies during the past five years. Normative data are available for samples of students from a variety of courses such as engineering, law, medicine, and arts and sciences. It is appropriate for any literate group.

The PWL is a self-report technique in which the respondent is instructed to mark those adjectives on the list which describe himself (either perceived, ideal, or real). Unless individual analysis is desired, best responses can be obtained by maintaining anonymity.

The score for self concept is obtained by counting the positive adjectives, counting the negative adjectives, and subtracting the latter from the former. Scores can be obtained for each of five dimensions—intellectual ability, character, social adjustment, emotional adjustment, and aesthetic ability—in this manner. Contradictions and omissions in self concept are identified by a division of the list into two parts consisting of antonyms.

The PWL is in printed form (in English) and may be obtained from the author. Although a price has not yet been set, some reasonable amount will be charged.

JOHN K. FISHER, Department of Psychology, Edinboro State College, Edinboro, Pennsylvania 16412

Self Concept as a Learner (Elementary Scale and Secondary Scale)

The purpose of this instrument is to assess a person's views of himself as a class member, a task-oriented individual, a problem solver, and a motivated individual. The instrument has been used with 200 upper-middle class white children in a suburban school system and with both elementary and secondary pupils in a culturally deprived community.

The elementary version is appropriate for use in grades three through six, and the secondary version for grades seven through twelve. There are no norms in terms of specific populations, but means and standard deviations from past uses are available. Results correlate fairly well with the *California Test of Personality*.

In culturally deprived areas the elementary scale should be read to the pupils, but this is not necessary for middle-class elementary or secondary pupils. In all cases it can be administered to groups.

Although there is a specific scoring procedure, this was not described.

Credit for development of the instrument belongs to Dr. Walter B. Waetjen (*see p. 158*). Fisher modified his (Waetjen's) secondary scale for use at the elementary level.

Permission for use of the scale must be granted by: Dr. Walter B. Waetjen, Vice-President, University of Maryland, College Park, Maryland.

For copies of the scales, write to John K. Fisher.

Copies of an elementary revision of Waetjen's Self Concept as a Learner are available from: Dr. Gordon Liddle, West Education Annex, University of Maryland, College Park, Maryland 20740.

ALAN F. FONTANA, Yale University, New Haven, Connecticut

The Measurement of Self-Esteem

The report of this instrument appeared in *Perceptual and Motor Skills* 23: 607-12; 1966. The instrument was used as a means of measuring self-esteem with graphic rating scales for 25 girls who participated in a three week sorority rushing program and who successfully attained membership.

A graphic rating scale accompanied by a general trait definition was used for each characteristic. Each scale was labeled from high to low and included trait adjectives specific to the characteristic under consideration. In order to counteract positional response set, the direction of the scales from high to low was randomly ordered.

Students made four types of judgments on each scale, thereby producing the following scores: (a) actual self-rating ("... the degree of the trait which you feel is characteristic of you"); (b) aspired self-rating ("the level you hope to attain; the point you feel you should strive to attain and have a realistic chance of reaching"); (c) worst self-rating ("you when you are having a bad day; the point that indicates your poorest performance on that trait"); and (d) an evaluation of each portion of the scale (division of each scale into highly desirable, acceptable, or undesirable locations for a rating of self). Weights of 2, 1, and 0 were assigned to the highly desirable, acceptable, and undesirable segments respectively. A total score for each rating of self was obtained by summing the values of all 18 scales. Thus, scores could range from 0-36 for each rating of self.

Odd-even and test-retest reliabilities indicate that aspired self-rating is not internally consistent, but both actual and worst self-ratings possess high internal consistency. Actual self-rating is significantly more stable than both the other self-ratings. Intercorrelations show that AsS is unrelated to AcS and WoS and that AcS and WoS are moderately related.

For a more extensive description of the instrument consult University Microfilms No. 64-12598.

JACK R. FRYMIER, College of Education, The Ohio State University, Columbus 43210

Faces Scale

This is an experimental scale designed to measure self concept and motivation of five- to ten-year-old children. In its present state, the instrument is for research purposes only. To date it has been used in two or three studies.

Forms A and B each contain 18 questions about the child's feelings toward family, school, friends, and self. After the teacher reads each question, the child responds by placing an "X" on either the smiling or the frowning face by the item number on his answer sheet. The Faces Scale may be administered to groups.

The teacher is provided with a sheet on which he is to rank boys and girls separately in the order that they have a positive self concept. This is an attempt to identify items which discriminate between youngsters with positive self concepts and those with less positive self concepts.

Normative data are not yet available. Item analyses have been executed in two studies using the instrument. A valid key has not been developed yet.

Examples of items:

How do you feel about how healthy and strong you are?

How do you feel about how much you know?

How do you feel about going to church?

How do you feel about the way your teacher treats you?

The Faces Scale may be reproduced or obtained from the author.

EUGENE L. GAIER, Faculty of Educational Studies, State University of New York, Foster Annex A, Buffalo 14214

Punishment Situation Index—PSI

This cartoon-type projective device was developed to assess characteristics of punishment responses in the mother-child relationship. In each picture, a child and his mother are depicted in a situation commonly followed by punishment, e.g., situations involving possible physical injury, unfavorable relationships with siblings, lying, and destruction of others' property. Spaces are provided above the figures as in comic strip cartoons for the subject to write in what he thinks each character is saying. Both mothers and children use the pictures,

but separate sets of pictures are used for boys and girls. The PSI yields four concepts operating in the punishment situation—from the child, his self concept (CC) and his concept of his mother (CM); from the mother, her self concept (MM) and her concept of the child (MC).

The obtained responses were scored using the three scoring factors, Extrapunitiveness, Intropunitiveness, and Impunitiveness, for direction of aggression developed by Rosenzweig.

Norms are available for a sample of boys and girls ranging in age from nine to twelve years from homes with professional fathers.

For explication of the instrument, refer to *Child Development*, Vol. 27, No. 4 and Vol. 28, No. 2 or contact the author.

MOHINDRA P. GILL, Department of Measurement and Evaluation,
The Ontario Institute for Studies in Education, 102 Bloor Street
West, Toronto 5, Ontario, Canada

The Self-Concept Scale

This instrument purports to measure the perceived self and the ideal self of high school students. It was used in a doctoral thesis which involved 1,424 ninth-grade students from five academic high schools in Toronto. Satisfactory reliability coefficients were obtained. Validity coefficients, using final average marks as the achievement criterion, were rather low—.42 (boys) and .35 (girls) for perceived self and .25 (boys) and .19 (girls) for ideal self.

This instrument is for group administration. The student rates himself on a four-point scale for each of 65 items in two forms. The inventory can be completed in about 35 minutes.

The instrument is not available at this time.

RICARDO GIRONA, 8B Hanna Hall, Bowling Green State University,
Bowling Green, Ohio 43402

Affect Scale

The Affect Scale, designed to assess level of positive self-regard, consists of 29 pairs of adjectives, e.g., ugly-beautiful, hostile-friendly, stingy-generous, which represent dimensions of self-regard. The subject is to rate himself along a seven-point scale between each pair. Scores on each scale range from one to seven, and total scores are computed by adding the scores of all the scales.

Normative data are not available. From a sample of 26 junior and sophomore college students, test-retest and split-half reliabilities

were found. Pilot studies have indicated that various populations will differ significantly.

The instrument is available on request for research purposes. Contact the author.

IRA J. GORDON, College of Education, University of Florida, Gainesville 33601

How I See Myself

This self-report instrument is designed to measure dimensions of self concept. It is available in a 40-item elementary form and a 42-item secondary form which have been used in a South American study and in several Florida studies.

Group administration is possible, and instructions and items may be read by the students or by the teacher to the students. Each item consists of two diametric statements with a five-point scale between them along which the student rates himself. Scores are obtained from the unweighted value of each item. Those items stated in such a way that the highest possible score indicates a feeling of inadequacy were transposed in order to have high scores represent a feeling of adequacy.

Sample items:

Elementary Form

Nothing gets me too mad.	1 2 3 4 5	I get mad easily and explode.
I don't stay with things and finish them.	1 2 3 4 5	I stay with something till I finish.

Secondary Form

I rarely get mad.	1 2 3 4 5	I get mad easily.
I have trouble staying with a job until I finish.	1 2 3 4 5	I stick with a job until I finish.

Using factorial analysis, 12 factors of self concept were identified in various combinations of grade levels and socioeconomic levels. Normative data were developed for six groups which were composed of third to twelfth grade white and Negro students. Reliability of the factors was about .80. There is no clear predictive validity as yet.

"How I See Myself" is available from the author although he feels that it needs further refinement. The price is not yet fixed.

Scales appear in the Florida Educational Research and Development Council's *Research Bulletin*, Vol. 3, No. 2, summer 1967.

EDMUND H. HENDERSON, The Reading Study Center, University of Delaware, Newark 19711

Self-Social Symbols Tasks (Preschool, Primary, and Adult Forms)

This is a nonverbal measure of self-social concepts. It has been used in several studies including cross-sectional and longitudinal developmental studies from preschool through high school and academic correlational studies of preschool through grade two. It is appropriate for all ages and nationalities, but normative data are not available, since this is an experimental instrument. Reliabilities have been reported satisfactory (median in the .80's), and substantial construct validity is claimed.

The 10-minute preschool form must be individually administered; primary and adult forms are group tests and require 20 to 30 minutes. All forms must be hand scored. Data may be treated by conventional statistics.

The instrument is available at cost from: Dr. Edmund H. Henderson, The Reading Study Center, University of Delaware, Newark, Delaware; and Dr. Barbara H. Long, Psychology Department, Goucher College, Baltimore, Maryland.

PHILIP S. HOLZMAN, The Menninger Foundation, Box 829, Topeka, Kansas 66601

Three Equivalent Forms of a Semantic Differential

This technique is *not* a test, but a method for assessing *changes* in attitude when one wants to control for the effects of prior judgments. Three equivalent forms were used to assess attitude toward one's own voice before hearing his own voice on tape, immediately after hearing his voice on tape, and again five minutes after the second administration. Results showed a reliable shift in attitude toward own voice immediately after listening to it; five minutes later the rating returned to prelistening levels. The technique is appropriate for men and women of average intelligence and above (ages 13 and above).

Seven-point scales are used for three subscores—activity, evaluation, and potency. Scores are obtained by averaging the ratings on the items in each subscale.

Information on validation of the technique is found in: Lolafaye Coyne and P. S. Holzman. "Three Equivalent Forms of a Semantic Differential Inventory." *Educational and Psychological Measurement* 26: 665-74; 1966. Reprints may be requested.

JOHN GOFF JONES, University of Texas Counseling Center, Austin 78712

Identity Development Rating Scale

Based primarily on Erik Erickson's theory, this instrument is designed to ascertain the degree of identity development acquired by the subject at his present age. It has been used with junior and senior high school students and will probably prove applicable to college level students.

Subjects are asked to respond to the question "Who am I?" without worrying about logic, order, or importance of statements. Five blank lines are presented to the subject to write his responses. Responses are then rated according to a five-point scale based on a continuum of identity ranging from diffuse and confused to well-developed. This technique of rating is the Identity Rating Scale.

Inter-rater reliability coefficients were .87 for males ($N = 167$) and .76 for females ($N = 150$).

For further information regarding the Identity Development Rating Scale, write to the author.

NADINE M. LAMBERT, Department of Education, University of California, Berkeley 94720

A Process for In-School Screening of Emotionally Handicapped Children

The ultimate purpose of the process is to provide a rapid, reliable, and economical method for identifying children with emotional handicaps. In design and purpose the process is similar to other screening activities carried on by schools in identifying children with health problems or vision or hearing loss. Administration of the instruments at planned intervals over a period of time can also be helpful in evaluating relationship patterns in individual children and in groups of children.

Applicable in all school grades, the process consists of three types of perceptions—teacher, peer, and self—which are scored and combined. The tests may be administered to groups by teachers without special training. Scoring does not require special training either.

At present, the process is to be used for research purposes only under the supervision of a competent psychologist or a specialist in the field of mental health. It is believed that the tests will be available for use under much less restrictive conditions within a period of several years. It is now available from: Office of Special Tests, Educational Testing Service, Princeton, New Jersey. (Three different instruments to assess self concept at primary, elementary, and secondary levels.)

GORDON P. LIDDLE, West Education Annex, University of Maryland,
College Park 20740

**Self Concept as a Learner—A Revision of
Walter B. Waetjen's Test (see p. 158)**

Waetjen's scale was revised to measure the self-image of children in grades three through six. This revised form was used with culturally disadvantaged third graders in a National Institute of Mental Health Study. Administration may be in groups with statements read aloud by the teacher or read by the students themselves.

The scale consists of 36 statements which pertain to four categories—motivation, intellectual ability, task orientation, and class membership. Students circle *yes* by statements they agree with and *no* by those they disagree with. One point is scored by each correct (as designated by the author) answer.

Examples of items:

I usually like to go to school.

I do well on tests.

I get my work done on time.

I find it hard to talk to classmates.

Normative data were established for a sample of 290 students.

The instrument is available. Contact the author.

MERLE L. MEACHAM, 318 Miller Hall, University of Washington,
Seattle 98105

Self-Concept Index of Motivation

Using a semantic differential technique, a seven-point scale for each of 15 pairs of adjectives related to motivation offers a measure of the self concept of motivation. The student is instructed to place a check on the scale between diametric adjectives such as "persevering" and "wavering" to indicate how he sees himself. Administration requires about 15 minutes.

This scale has been used only with junior college populations, and no normative data are available. A complete item analysis is in the author's thesis (on file at Washington State University, Pullman).

If contacted, the author will share further information.

MICHIGAN STATE UNIVERSITY, Bureau of Educational Research,
East Lansing

Self Concept of Ability—General and Specific

This self-rating scale in two forms, A (general) and B (specific), consists of eight questions related to school ability. The eight questions

are the same in both forms, but the answer formats are different. In Form A the subject rates himself on a five-point scale in answer to each question; in Form B the subject rates himself four times, in regard to mathematics, to English, to social studies, and to science, for each question.

OKLAHOMA CITY, OKLAHOMA—Federal Program

Children's Self-Concept Scale

The instrument consists of 100 simple declarative statements with Likert-type scoring. Although not stated, the age level for which the instrument is appropriate appears to be as low as seven or eight, provided the items are read to the students. The vocabulary is very simple.

Sample items from the scale are:

- 20. Sometimes I cannot do anything right.
- 30. If I could, I would hurt my friends.
- 50. People really like me.
- 80. Sometimes my friends try to hurt me.

JAMES PARKER, Department of Education, Georgia Southern College, Statesboro

About Me

This self-report instrument assesses five areas of self concept which could be expected to be expressed in behavior in the school environment. These five areas are the self, the self in relation to others, the self as achieving, the self in school, and the physical self. There are six items for each of the five areas.

Each of the 30 items consists of a positive and a negative statement at opposite ends of a continuum. The respondent is to rate himself along a five-point scale between the two statements. The following are sample items taken from the instrument:

	1	2	3	4	5	
I'm good in school work.						I'm not good in school work.
I'm popular.						I'm not too popular.
I don't get tired quickly.						I get tired quickly.
I'm not tall enough.						I'm tall enough.
I'm proud of me.						I'm not too proud of me.

About Me was constructed for use in a dissertation study. It was used with 60 grade pupils and is appropriate for students in grades

four through six. Individual or group administration is possible. No rigorous normative or statistical data are available. Scores are derived by summing the numerical values of individual items. High scores indicate a negative self concept; low scores, a positive self concept.

The author will furnish the instrument upon receipt of return postage.

ERNEST L. PETERS, Director, Division of Cooperative Research Services, Department of Public Instruction, Box 911, Harrisburg, Pennsylvania 17126

Self Concept as a Driver Scale

This scale consists of 115 statements which the subject applies to himself and responds on a five-point scale which ranges from false to true. All statements refer to actions and attitudes regarding the operation of a vehicle.

Illustrative items:

- I try not to take chances when I am driving.
- I get tensed up when the car stalls in traffic.
- I think my reaction time is good.
- I enjoy the thrill of driving fast.

ELLEN V. PIERS and DALE B. HARRIS, The Pennsylvania State University, 177 Borrowes Building, University Park 16802

Piers-Harris Self-Concept Scale

This instrument consists of 80 declarative statements for which the subject responds "yes" or "no" to indicate whether or not they apply to him. Through factor analysis the following six major dimensions were identified: behavior, general and academic status, physical appearance and attributes, anxiety, popularity, and happiness and satisfaction. This scale is appropriate for students in the third grade and above. In grades three, four, five, and six the statements should be read to the students; only in the seventh grade and above should students be left to read to themselves.

Some of the items from the scale are:

- 1. My classmates make fun of me.
- 16. I have good ideas.
- 31. In school I am a dreamer.
- 46. I am among the last to be chosen for games.
- 61. When I try to make something, everything seems to go wrong.
- 76. I cry easily.

Items are scored in the direction of high (adequate) self concept. It is suggested that the total number of "highs" be added, the total number of lows added and written below the "highs." A key for scoring is supplied.

Contact the authors for further information.

WALTER RECKLESS, Department of Sociology and Anthropology,
Ohio State University, Columbus 43210

**"The Way It Looks to Me"—The O.S.U. Delinquency Project's
Self-Concept Instrument**

This inventory is not a hard measure for assessing individuals; it gives a directional indication for criterion groups, e.g., the better students in a class *versus* the poorer students. Significant differences in the mean scores between criterion groups, not between two or more individuals, should be sought. It is appropriate for use with 12- to 14-year-olds.

The Self-Concept Inventory consists of questions read aloud to the students by the examiner. The following are examples of the questions:

"Do you think that things are pretty well stacked against you?"

"Will you probably be taken to juvenile court sometime?"

"Did anyone ever tell you that you have a problem?"

The students respond to statements by circling Y (yes), N (no), or DK (don't know). These answers are scored on a three-point scale. Total score is the sum of the individual item scores. Two forms, a long one of 32 items (16 significant and 16 filler items) and a short form of 14 items (7 significant and 7 filler items), are available. Only the significant items are scored, and high scores are in the unfavorable direction. Normative and statistical data were not specified.

Permission may be requested by any professional worker to use the OSUDP's Self-Concept Instrument.

MILDRED T. RICHARDSON, The Devereux Foundation, Education
Center, Devon, Pennsylvania

**Discrepancy Measurement Relating Student Self Concept of
Mental Ability with Mental Health Stability**

This technique was designed to provide a prediction of a student's relationship between (a) discrepancy measures relating a student's self concept of his mental ability, and (b) his attitudes towards various aspects of the life situation. It incorporates the Beier Sentence Com-

pletion Test and has been used as a screening device for counseling priority. It is appropriate for the secondary level school population.

As this technique was developed for a doctoral dissertation, data regarding validity and reliability are very local in nature but show promise. Further information may be obtained from the above address, and a copy of the technique can be procured for the cost of duplication.

JOHN E. RILEY, Drawer E—TWU Station, Texas Woman's University, Denton 76204

Animal Picture Q Sort

This is a technique for assessing the self concepts of elementary and preschool children. It is particularly useful in determining sense of adequacy in children's sex roles. The author feels that in its present form it is of little use to anyone, but he plans to continue work on it this year.

Pupils sort 36 animal pictures into a seven-category, forced normal distribution from "like me" to "unlike me." Each sort is tested for homogeneity of variance using an F test; then each sort receives an analysis of variance for a balanced block design of two factors (male adequacy and female adequacy) at three levels each (high, medium, and low).

There are no normative data since the design was built around Stephenson's Q technique which is ipsetive rather than normative. Data are available, however, on about 120 second- and third-grade boys from two schools. Indications of reliability were suggested by an F max test for homogeneity of variance for each individual sort. The instrument discriminated between boys who were rated differently in masculine behavior. The statistical procedures employed in Riley's dissertation make this technique very cumbersome.

"Animal Picture Q Sort" is adapted from an animal game for a projective technique. The game is available through: Milton Bradley and Company, Springfield, Massachusetts. (Animal Lotto—\$1.00.)

LOIS STILLWELL, 3921 Woodthrust Road, Akron, Ohio 44314

Specific and Global Self Concept (a derivative of Osgood's Semantic Differential)

The instrument purports to measure specific as well as global self concept, e.g., "myself as a student," "myself as a reader." It could be used to assess attitudes toward other things if one wanted a comparison between self concept and those attitudes. The technique is appropriate for children in grades four and above and, with revision and simplification of terms, for primary grades.

It consists of nine scales, each of which presents five steps along a continuum between two diametric adjectives, e.g., very strong, somewhat strong, average, somewhat weak, very weak. Possible score on each scale ranges from one to five. Total score is the sum of scores on all nine scales.

Reliability ranged from .55 to .90 for girls and from .63 to .85 for boys. Correlations between various aspects of self concept indicate that construct validity is at an adequate level.

Copies of the scale are available from the author. It is mimeographed with the following directions at the top: "Circle the term in each row which best describes _____."

R. WRAY STROWIG, Department of Counseling and Behavioral Studies, University of Wisconsin, Madison 53706

Student Self-Expectations Inventory—SSE

This instrument was designed to measure the expectations individuals hold for themselves as students. It was developed and used with rural ninth- and twelfth-grade students. There are no normative data other than means and variances for the samples mentioned above, but the author suggests that the inventory is probably appropriate for urban youth also.

Practically self-administering, the test requires about 10 minutes examination time. It yields one score; high scores are indicative of high standards.

The following are samples of items taken from the SSE:

As a student, I expect myself to:

1. do good work even in classes I don't like.
4. attend school regularly.
11. listen carefully to class discussions.
25. take school seriously.
26. be hard to get to know.

A student responds to items in terms of the extent to which he feels that each would apply to himself. Answers are keyed according to the degree of agreement between the respondents' answers and responses of a sample of students known to be very successful academically (high GPA).

The instrument is copyrighted, and the author will permit its use only for research purposes with the understanding that detailed results will be shared with him.

WALTER B. WAETJEN, Vice-President, University of Maryland, College Park

Self Concept as a Learner Scale—SCAL

The SCAL is divided into four components which constitute certain dimensions of one's self concept as a learner. Items within each component are judged in terms of the way an adequate learner would respond and may be either positive or negative statements. The four components are motivation, task orientation, problem-solving or intellectual ability, and class membership.

Examples of items:

I am usually eager to go to class.

I do well when I work alone.

I can't express my ideas in writing very well.

I find it hard to talk with classmates.

There is a total of 50 statements. Students respond to statements using categories of completely false, mostly false, partly true and partly false, mostly true, and completely true. For positive statements the categories are scored one to five respectively, while for negative statements this procedure is reversed.

Contact the author for permission to use the scale.

(See also adaptations by Gordon Liddle on page 152 and by John Fisher on page 145.)

Indexes to the Inventory

By Authors

- Acker, Mary, 97
 Agazarian, Yvonne, 104-105
 Alcorn, John D., 99
 Amatora, S. Mary, 129
 Amidon, Edmund, 99-100
 Anderson, Harry E., 141
 Aronoff, Joel, 116
 Aspy, David N., 100
 Astin, Helen S., 107

 Barberousse, Eleanor H., 95, 100-101
 Barker, Donald G., 90, 116
 Behring, Daniel W., 117
 Bemis, Katherine, 130
 Bentley, Ralph, 90-91
 Bills, Robert E., 107
 Blackburn, Joe M., 117
 Bloch, Donald A., 130
 Borgotta, Edgar F., 130-31
 Brown, Oliver H., 118

 Campbell, Paul B., 118
 Campos, Leonard P., 132
 Carney, Richard E., 118-19
 Cattell, Raymond B., 132-34, 140
 Cincinnati Public Schools, Division of
 Psychological Services and Division
 of Program Development, Cincin-
 nati, Ohio, 143
 Coan, Richard W., 134
 Collier County Board of Public In-
 struction, Collier County, Naples,
 Florida, 119
 Combs, Charles F., 107
 Coopersmith, Stanley, 143
 Courson, Clifford C., 144
 Cunningham, Claude D., 119-20

 Dean, Dwight G., 134-35
 Dedrick, Charles Van Loan, 144
 Denny, David A., 95-96
 Denny, Terry, 141
 Deo, Pratibha, 91, 96, 108, 145
 Duncan, James, 101

 Eber, H. W., 132-33
 Eysenck, H. J., 135

 Farquhar, William W., 120
 Fishburn, William R., 108-109
 Fisher, John K., 145-46
 Flanders, Ned A., 101
 Fontana, Alan F., 146-47
 French, John W., 120-21
 Frymier, Jack R., 96-97, 101, 121, 147
 Furneaux, W. D., 136-37

 Gaier, Eugene L., 147-48
 Gibson, H. B., 136-37
 Gilbert, Albin R., 137
 Gill, Mohindra P., 148
 Girona, Ricardo, 148-49
 Gordon, Ira J., 149
 Gulo, E. Vaughn, 121-22

 Hall, Nason, 122
 Harris, Dale B., 109, 154-55
 Hayes, Robert B., 109-10
 Henderson, Edmund H., 150
 Holliman, Neil B., 123
 Holzman, Philip S., 150
 Honigman, Fred K., 102

 Jones, John Goff, 151
 Jordan, John E., 92, 123

 Kerlinger, Fred N., 92
 Kline, Paul, 138

 Lambert, Nadine M., 151
 Lemeshow, Seymour, 110
 Liberty, Paul G., 102-103, 138
 Liddle, Gordon P., 152
 Lowery, Lawrence F., 123-24

 McCandless, Boyd R., 124
 McReynolds, Paul, 97

 Mason, Russell E., 139
 Maw, Ethel W., 125
 Maw, Wallace H., 125
 Meacham, Merle L., 152
 Medley, Donald M., 103
 Michigan State University Bureau of
 Educational Research, East Lan-
 sing, 152-53
 Mooney, Ross L., 110-11

- Nelson, Paul A., 125-26
 Novick, Jack, 111
 Oklahoma City, Oklahoma—Federal Program, 153
 Ong, Jin, 111
 Parker, James, 153-54
 Pervin, Lawrence A., 104
 Peters, Ernest L., 112, 154
 Piers, Ellen V., 154-55
 Plutchik, Robert, 139
 Porter, Rutherford B., 140
 Pumroy, Donald K., 92-93
 Reckless, Walter, 155
 Reeback, Robert T., 112
 Rempel, Avesno M., 90-91
 Richardson, Mildred T., 155-56
 Riley, John E., 156
 Rookey, Thomas J., 97-98
 Rosenberg, B. G., 113-14
 Rowe, William Frank, 126
 San Francisco State College, San Francisco, California, 141-42
 Scheier, Ivan H., 133, 140
 Schlessner, George E., 126
 Simon, Anita, 104-105
 Spaulding, Robert L., 105, 113
 Stewart, Lawrence H., 113
 Stillwell, Lois, 156-57
 Stollah, Gary E., 140
 Strowig, R. Wray, 157
 Sutton-Smith, 113-114
 Thomas, Walter L., 114
 Thompson, Glen Robbins, 126-27
 Tobiasson, Ray, 93
 Torrance, E. Paul, 98
 Van Looy, Johanna C., 115
 Waetjen, Walter B., 158
 Wiesen, Henry H., 106
 Wood, Frank H., 127-28
 Wright, Benjamin D., 93-94
 Wright, E. N., 106
 Wrightsman, Lawrence S., 94, 128
 Zinker, Joseph C., 128

By Titles of Measures

- About Me, 153-54
 Achievement-Oriented Scale, 118-19
 Activities Preference Achievement Scale (APAS), 117
 Affect Scale, 148-49
 Ai 3, 138
 Alienation Scale, 154
 An Attitude Scale for Punishment, 91
 An Attitude Scale for Ragging, 91
 Animal Picture Q Sort, 156
 Attitude Scales, 90-94
 Attitudes Toward Education, 123
 Attitudes Toward Handicapping Conditions, 92
 Attitudes Toward Professors, 121-22
 Attitudes Toward Riding the School Bus, 90
 Behavioral Maturity Scale, 141
 Behavioral Self-Rating Form (BSR), 130-31
 Children's Personality Questionnaire (CPQ), 140
 Children's Self-Concept Scale, 153
 Combs School Apperception Test, 107
 Content Attitude Test, 125-26
 Coping Analysis Schedule for Educational Settings (CASES), 105
 Creativity, 95-98
 Cross-Cultural Functional Personality Analysis Inventory, 139
 D-I Inventory, 108
 Denny-Ives Creativity Test, 95-96
 Deviant Behavior Inventory (DBI), 111
 Deviant Behavior Inventory for Children, 130
 Differential Value Profile, 114
 Discrepancy Measurement Relating Student Self Concept of Mental Ability with Mental Health Stability, 155-56
 Dissatisfaction Magnitude Scale (DIMS), 93
 Draw-A-Classroom Test, 106
 Duncan Teaching Situation Reaction Test (TSRT), 101

- Early School Personality Questionnaire (ESPQ), 134
Education Scale VII (ES-VII), 92
Emotional Maturity Scale, 134-35
Emotions Profile Index (E.P.I.), 139
Experimental Procedure for Measuring Reading Achievement Motivation in Children, 127-28
- Faces Scale, 147
- The Gibson Spiral Maze, 136
Goodenough-Harris Drawing Test, 109
Group Counseling Evaluation Scale (Form II), 108-109
- Hayes Pupil-Teacher Reaction Scale, 109-10
High School Personality Questionnaire (HSPQ), 132
"How I Feel" Attitude Inventory Test, 119
How I See Myself, 149
- IPAT Anxiety Scale Questionnaire, 133
Identity Development Rating Scale, 151
Index of Adjustment and Values (IAV), 107
Inference, 144
Intensity of Involvement Scale (Observation), 124
Interaction, 99-106
Interaction Analysis, 101
Interest Assessment Scales, 113
The Interpersonal Orientation Scale (IOS), 99
- Junior Eysenck Personality Inventory, 135
The Junior High Boy, 122
Junior High School Articulation Scale, 117
Junior Index of Motivation—JIM Scale, 121
Junior Maudsley Inventory, see: The New Junior Maudsley Inventory, 136-37
- Latency-weighted Personality Testing (Technique), 137
Levine-Elzey Preschool Social Competency Scale, 141-42
- M-Scales, 120
Maryland Parent Attitude Survey (MPAS), 92-93
The Measurement of Self-Esteem, 146-47
Miscellaneous, 107-15
Mooney Problem Checklist, 110-11
Motivation, 116-28
Multidimensional Analysis of Classroom Interaction (MACI), 102
- Neuroticism Scale Questionnaire (NSQ), 140
The New Junior Maudsley Inventory, 136-37
- Obscure Figures Test, 97
Observation Schedule and Record—OSCAR 2a, 103; OSCAR 5V, 103
Ohio State Picture Preference Scale (OSPPS), 96-97
The Opposite-Form Procedure in Inventory Construction and Research, 111
Organizational Climate in the Classroom (OCIC), 106
- Paired Comparison Technique, 123
Pennsylvania Assessment of Creative Tendency, 97-98
Pennsylvania Citizenship Assessment Instrument—Fifth Grade, 112
Perceptual Score Sheet, 144
Personal Values Inventory, 126
Personality, 129-40
Personality Rating Scale, 129
Personality Word List, 145
Philosophies of Human Nature Scale, 94
A Picture Test for Social Distance, 108
Piers-Harris Self-Concept Scale, 154-55
Preschool Academic Sentiment Scale (PASS), 126-27
Problem List, 140
A Process for In-School Screening of Emotionally Handicapped Children, 151
The Projective Tests of Attitudes (PTOA), 123-24
Punishment Situation Index-PSI, 147-48

- Pupil Creativity Concept Q-Sort, 95
The Purdue Teacher Opinionnaire (PTO), 90-91
- Q-Sort for the Hierarchy of Needs, 128
Questionnaire on Motivation in College, 120-21
- Readiness*, 141-42
Reading Attitude Inventory, 118
Reading Percepts Interview Schedule, 141
- SWCEL Classroom Observer Rating Schedule, 102-103
SWCEL Student Questionnaire, 138
Scale of Attitudes Toward School Guidance, 116
Scales for School and Law Attitudes, 122
School Attitude Q-Sort, 126
School Morale Scale, 128
Self Concept, 143-58
Self Concept as a Driver Scale, 154
Self Concept as a Learner—A Revision of Walter B. Waetjen's Test, 152
Self Concept as a Learner (Elementary and Secondary Scale), 145-46
Self Concept as a Learner Scale—SCAL, 158
Self-Concept Index of Motivation, 152
Self Concept of Ability—General and Specific, 152-53
The Self-Concept Scale, 148
Self-Esteem Inventory, 143
Self-Report Inventory—Form R-3, 118
Self-Social Symbols Tasks (Preschool, Primary, and Adult Forms), 150
Sentence Completion Test, 116
Sequential Analysis of Verbal Interaction (SAVI), 104-105
- A Short Test of Personality: The S-ident Form, 131
Situational Test of Empathy, 107
The Sixteen Personality Factor Test (16PF) Forms A, B, C, 132-33
Sociometric Reputation Nomination Scale, 100-101
The Spaulding Teacher Activity Rating Schedule (STARS), 113
Specific and Global Self Concept, 156-57
Story Completion Technique—Children's Form, 132
Student Self-Expectations Inventory—SSE, 157
- Teacher Observation Personality Schedule (TOPS), 130
Teacher Operational Problems Identification, 110
Teaching Attitudes Questionnaire, 1962, 93-94
Test Attitude Scale, 119-20
Tests of Creativity, 96
Three Equivalent Forms of a Semantic Differential, 150
Torrance Tests of Creative Thinking, 98
Transactional Analysis of Personality and Environment (TAPE), 104
Truax Scales for Empathy, Congruence, and Positive Regard, 100
Van Looy's Expectancy Scale, 115
The Verbal Interaction Category System (VICS), 99-100
The Vigilance Game, 112
"The Way It Looks to Me"—The O.S.U. Delinquency Project's Self-Concept Instrument, 155
What I Am Like, 143
What I Like To Do (An Impulsivity Scale), 113-14
The You Test, 125

By Abbreviations Associated with Titles

- Ai 3, Measure of Freudian Anal Character, 138
AO, Achievement-Orientation Scale, 118-19
APAS, Activities Preference Achievement Scale, 117
- BSR, Behavioral Self-Rating Form, 130-31
CASES, Coping Analysis Schedule for Educational Settings, 105
CPQ, Children's Personality Questionnaire, 140

- DBI, Deviant Behavior Inventory, 111
 DIMS, Dissatisfaction Magnitude Scale 93
 DVP, Differential Value Profile, 114
 E.P.I., Emotions Profile Index, 139
 ES-VII, Education Scale VII, 92
 ESPQ, Early School Personality Questionnaire, 134
 HSPQ, High School Personality Questionnaire, 132
 IAV, Index of Adjustment and Values, 107
 IOS, The Interpersonal Orientation Scale, 99
 IPAT Anxiety Scale Questionnaire, 133
 JIM-Scale, Junior Index of Motivation, 121
 JMPI, Junior Maudsley Personality Inventory, 136-37
 MACI, Multidimensional Analysis of Classroom Interaction, 102
 MPAS, Maryland Parent Attitude Survey, 92-93
 NSQ, Neuroticism Scale Questionnaire, 140
 OCIC, Organizational Climate in the Classroom, 106
 OFT, Obscure Figures Test, 97
 OScAR, Observation Schedule and Record 2a, 103; 5v, 103
 OSPPS, Ohio State Picture Preference Scale, 96-97
 PACT, Pennsylvania Assessment of Creative Tendency, 97-98
 PASS, Preschool Academic Sentiment Scale, 126-27
 PSI, Punishment Situation Index, 147-48
 PTO, The Purdue Teacher Opinionnaire, 90-91
 PTOA, The Projective Tests of Attitudes, 123-24
 PWL, Personality Word List, 145
 SAVI, Sequential Analysis of Verbal Interaction, 104-105
 SCAL, Self Concept as a Learner Scale, 158
 S-ident Form, A Short Test of Personality, 131
 16PF, The Sixteen Personality Factor Test, Forms A, B, C, 132-33
 SSE, Student Self-Expectations Inventory, 157
 STARS, The Spaulding Teacher Activity Rating Schedule, 113
 STE, Situational Test of Empathy, 107
 SWCEL Classroom Observer Rating Schedule, 102-103
 SWCEL Student Questionnaire, 138
 TAPE, Transactional Analysis of Personality and Environment, 104
 TOPS, Teacher Observation Personality Schedule, 130
 TSRT, Duncan Teaching Situation Reaction Test, 101
 VICS, The Verbal Interaction Category System, 99-100

Contributors

WALCOTT H. BEATTY

Professor, Department of Psychology, San Francisco State College, San Francisco, California

DONALD J. DOWD

President of Island Enterprises, Ltd., Marathon, Florida; formerly Economics Coordinator, Dade County Public Schools, Florida

ROBERT E. STAKE

Professor of Psychology, University of Illinois, Urbana

DANIEL L. STUFFLEBEAM

Director, Evaluation Center, The Ohio State University, Columbus

RALPH W. TYLER

Director Emeritus, Center for Advanced Study in the Behavioral Sciences, Stanford, California

SARA C. WEST

Graduate student, The Ohio State University, Columbus; formerly Research Associate, Dade County Public Schools, Florida

Members of the ASCD Council on Assessment of Educational Outcomes

WALCOTT H. BEATTY, *Chairman*

DONALD J. DOWD

JACK R. FRYMIER

Professor, College of Education, The Ohio State University, Columbus

EARL C. KELLEY

Distinguished Professor Emeritus of Education, Wayne State University, Detroit, Michigan (*Council member, 1965-67*)

ASCD Publications

(The NEA stock number appears in parentheses after each title.)

Yearbooks

Balance in the Curriculum (610-17274)	\$4.00
Evaluation as Feedback and Guide (610-17700)	\$6.50
Fostering Mental Health in Our Schools (610-17256)	\$3.00
Guidance in the Curriculum (610-17266)	\$3.75
Individualizing Instruction (610-17264)	\$4.00
Leadership for Improving Instruction (610-17454)	\$3.75
Learning and Mental Health in the School (610-17674)	\$5.00
Learning and the Teacher (610-17270)	\$3.75
Life Skills in School and Society (610-17786)	\$5.50
New Insights and the Curriculum (610-17548)	\$5.00
Perceiving, Behaving, Becoming: A New Focus for Education (610-17278)	\$4.50
Research for Curriculum Improvement (610-17268)	\$4.00
Role of Supervisor and Curriculum Director (610-17624)	\$4.50
Youth Education: Problems, Perspectives, Promises (610-17746)	\$5.50

Booklets

Assessing and Using Curriculum Content (611-17662)	\$1.00
Better Than Rating (611-17298)	\$1.25
Changing Curriculum Content (611-17600)	\$1.00
The Changing Curriculum: Mathematics (611-17724)	\$2.00
The Changing Curriculum: Modern Foreign Languages (611-17764)	\$2.00
The Changing Curriculum: Science (611-17704)	\$1.50
Changing Supervision for Changing Times (611-17802)	\$2.00
Children's Social Learning (611-17326)	\$1.75
Collective Negotiation in Curriculum and Instruction (611-17728)	\$1.00
Criteria for Theories of Instruction (611-17756)	\$2.00
Curriculum Change: Direction and Process (611-17698)	\$2.00
Curriculum Decisions ↔ Social Realities (611-17770)	\$2.75
A Curriculum for Children (611-17790)	\$2.75
Curriculum Materials 1969 (611-17784)	\$2.00
Discipline for Today's Children and Youth (611-17314)	\$1.00
Early Childhood Education Today (611-17766)	\$2.00
Educating the Children of the Poor (611-17762)	\$2.00

Elementary School Mathematics: A Guide to Current Research (611-17752)	\$2.75
Elementary School Science: A Guide to Current Research (611-17726)	\$2.25
The Elementary School We Need (611-17636)	\$1.25
Extending the School Year (611-17340)	\$1.25
Freeing Capacity To Learn (611-17322)	\$1.00
Guidelines for Elementary Social Studies (611-17738)	\$1.50
The High School We Need (611-17312)	\$1.50
Human Variability and Learning (611-17332)	\$1.50
The Humanities and the Curriculum (611-17708)	\$2.00
Humanizing Education: The Person in the Process (611-17722)	\$2.25
Humanizing the Secondary School (611-17780)	\$2.75
Improving Educational Assessment & An Inventory of Measures of Affective Behavior (611-17804)	\$3.00
Improving Language Arts Instruction Through Research (611-17560)	\$2.75
Influences in Curriculum Change (611-17730)	\$2.25
Intellectual Development: Another Look (611-17618)	\$1.75
The Junior High School We Need (611-17338)	\$1.00
The Junior High School We Saw (611-17604)	\$1.50
Juvenile Delinquency (611-17306)	\$1.00
Language and Meaning (611-17696)	\$2.75
Learning More About Learning (611-17310)	\$1.00
Linguistics and the Classroom Teacher (611-17720)	\$2.75
New Curriculum Developments (611-17634)	\$1.75
New Dimensions in Learning (611-17336)	\$1.50
The New Elementary School (611-17734)	\$2.50
Nurturing Individual Potential (611-17606)	\$1.50
Personalized Supervision (611-17680)	\$1.75
Strategy for Curriculum Change (611-17666)	\$1.25
Supervision in Action (611-17346)	\$1.25
Supervision: Emerging Profession (611-17796)	\$5.00
Supervision: Perspectives and Propositions (611-17732)	\$2.00
The Supervisor: Agent for Change in Teaching (611-17702)	\$3.25
The Supervisor: New Demands, New Dimensions (611-17782)	\$2.50
The Supervisor's Role in Negotiation (611-17798)	\$1.75
Theories of Instruction (611-17668)	\$2.00
Toward Professional Maturity (611-17740)	\$1.50
What Are the Sources of the Curriculum? (611-17522)	\$1.50
Child Growth Chart (618-17442)	\$1.25

Discounts on quantity orders of same title to single address: 2-9 copies, 10%; 10 or more copies, 20%. Orders for \$2 or less must be accompanied by remittance. Postage and handling will be charged on all orders not accompanied by payment. The NEA stock number of each publication must be listed when ordering.

Subscription to Educational Leadership—\$6.50 a year. ASCD Membership dues: regular (subscription and yearbook)—\$20.00 a year; Comprehensive (includes subscription and yearbook plus other publications issued during period of the membership)—\$30.00 a year.

Order from: Association for Supervision and Curriculum Development, NEA
1201 Sixteenth Street, N.W. Washington, D.C. 20036